NOTE TO USERS

The original manuscript received by UMI contains pages with slanted print. Pages were microfilmed as received.

This reproduction is the best copy available

UMI
EVALUATION OF
PAEDIATRIC PHYSICAL DISABILITY
&
EXPLORATION OF
CONTRIBUTING FACTORS

BY
NANCY LYNN YOUNG

A thesis submitted in conformity with the requirements for the degree of
Doctor of Philosophy
Institute of Medical Science
University of Toronto

© Copyright by Nancy Lynn Young 1997
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
Evaluation of Paediatric Physical Disability & Exploration of Contributing Factors

Nancy Lynn Young

Doctor of Philosophy, Institute of Medical Science, University of Toronto, 1997

ABSTRACT

Introduction: The purpose of this research was to assess the measurement properties of the Activities Scale for Kids or ASK (a physical disability measure for five to 15 year old children with musculoskeletal conditions), and to explore the magnitude and nature of the difference between measures of what children can do (capability) and what they usually do (performance).

Methods: The ASK's measurement properties were assessed by: item characteristic analysis, comparison of the ASK to clinician observation, clinician global ratings, the Childhood Health Assessment Questionnaire (CHAQ), the construct of clinically meaningful change, and the Health Utilities Index (HUI3). The difference between capability and performance were explored by quantitative and qualitative comparisons of ASK-capability and ASK-performance reports.

Results: Item characteristic analysis confirmed the unidimensionality and item structure of the ASK (n=200, mean age 10 years). Validity was confirmed by strong relationships with clinician assessment (ICC=0.89, lower limit of confidence interval 0.79), clinician global ratings (r<0.0001), the CHAQ (r=0.81, CI 0.76-0.86), clinical change (r= 0.77, CI 0.58-0.88), and related HUI3 sub-scales (r=0.74 and r=0.10) and minimal relationships with unrelated HUI3 sub-scales (r= -0.15 and r=0.09). Comparison of performance and capability scores confirmed that there was a difference (p<0.0001) related to the framing of disability questions. The interviews with respondents identified several factors related to the expression
of disability: affinity for the activity, formidability of the activity, and the availability of modifications (i.e., assistive devices).

Conclusions: The ASK has acceptable item characteristics and is a sensible, reliable and valid measure of physical disability in children. Since capability scores significantly exceed performance scores, the situation of relevance must be carefully considered in selecting an outcome measure. The factors that contribute to this difference may offer insight into alternative mechanisms by which to modify physical disability in this group of children.
ACKNOWLEDGMENTS

This research was done under the careful guidance of Dr. J. Ivan Williams (thesis supervisor). His broad vision and consistent facilitation provided a superb environment for the growth of my knowledge and generation of this thesis. His remarkable dedication and sense of perspective were greatly appreciated, particularly during more challenging components of the writing and revision process. Dr. James G. Wright (thesis committee member) provided consistent support, and was most appreciated for his expertise in paediatric orthopaedics and measurement theory. Dr. Karen Yoshida (thesis committee member) provided a rehabilitation perspective and contributed qualitative research experience to the project. My sincerest thanks to these individuals, without whom this research would not have been possible.

I also would like to thank the Ontario Ministry of Health for providing fellowship funding, the Easter Seal Research Institute for providing the operating grant, and the many clinicians at the Bloorview MacMillan Centre, Children’s Hospital of Michigan, Erinoak, the Hospital for Sick Children and Variety Village who assisted in the recruitment of children for this research.

The children and parents who assisted in this research are equally worthy of recognition for their contributions. These families provided the initial inspiration for the research and a wealth of very valuable information. I hope that this research will provide sufficient new knowledge, the practical measurement methods needed to evaluate treatment interventions, and thus help minimize the expression of disability among these children and their peers in the future.

Last, but not least, I thank my husband Stephen. His unwavering support, patience, humor and musical interludes fortified me during this quest.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter 1</th>
<th>Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Historical Background</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Statement of the Problem</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Summary of Purpose</td>
<td>9</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>Review of Literature</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Conceptual Framework &amp; Definition of Disability</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Methods of Scale Development</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Existing Measures</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Overview of the ASK</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Role of The Environment in Disability Evaluation</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Proposed Work</td>
<td>47</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Methods</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Overview of Methods</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Part A: ASK Measurement Properties</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Item Characteristic Analysis</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Validation Studies</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Part B: Exploring the Determinants of Physical Disability</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Quantitative comparison of Performance and Capability</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Qualitative Comparison of Performance and Capability</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Summary of Methods</td>
<td>80</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Results Part A</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>ASK Measurement Properties</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>Descriptive Statistics</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Item Characteristic Analysis</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>Validation Studies</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Summary of Measurement Property Results</td>
<td>107</td>
</tr>
</tbody>
</table>
Chapter 5 Results Part B ......................................................... 108
  Exploring the Determinants of Physical Disability ......................... 108
  Quantitative Comparison of ASKp and ASKc ................................ 108
  Qualitative Comparison of Performance and Capability ....................... 111
  Summary of Exploration of Determinants Results .......................... 162

Chapter 6 Discussion & Conclusions ............................................. 165

References .................................................................................... 189

Appendices .................................................................................... 208
  Appendix I Alternate Disablement Models .................................... 208
  Appendix II Background on Rasch Modeling .................................. 211
  Appendix III Item History ............................................................ 226
  Appendix IV ASK Development Methods ..................................... 241
  Appendix V Ethics Approvals & Consent forms .............................. 242
  Appendix VI Global Rating Algorithm ........................................ 251
  Appendix VII ASK Questionnaires .............................................. 252
  Appendix VIII Heath Utilities Index Mark III Questionnaire .............. 272
  Appendix IX Scale Aggregation Formulae ..................................... 273
  Appendix X Item Logits in Questionnaire Order ............................ 274
LIST OF TABLES

Table 1  The Five Purposes of Measurement ................................................. 23
Table 2  Feinstein’s Criteria for Appraising Measures ................................. 25
Table 3  Review of Existing Scales ................................................................. 32
Table 4  Summary of Methods ........................................................................ 49
Table 5  ASK Domains and Sample Items ....................................................... 59
Table 6  Breakdown of Recruitment & Completion ........................................... 83
Table 7  Distribution of Sample Characteristics ............................................ 84
Table 8  Item Response Frequencies for ASKp From 200 Children ................. 87
Table 9  Number of Items in Children’s ASKp Summary Scores ...................... 90
Table 10 ASKp53 Rasch Model ................................................................ ...... 93
Table 11 Pearson’s Correlation Matrix Between ASK Versions ...................... 96
Table 12 Summary of Item Validity and Item Reliability .................................. 97
Table 13 ASK Scores by Clinicians’ Global Ratings of Disability Severity ........... 99
Table 14 Convergent and Divergent Spearman’s Correlation Coefficients ......... 101
Table 15 Frequency of Ceiling and Floor Effects by Scale .............................. 102
Table 16 Ceiling Effects in a Selected Subset ............................................... 103
Table 17 Responsiveness Results by Prognosis ............................................. 105
Table 18 Summary of Results Part A ............................................................... 107
Table 19 Demographic Distribution of Interview Participants ....................... 112
Table 20 Distribution of Disability Scores in Qualitative Sample .................... 113
Table 21 Summary of Results Part B ............................................................... 164

Appendix Tables

Table A1  Preliminary Item Reduction Data and Results ............................... 228
Table A2  Individual Item Responsiveness ..................................................... 230
Table A3  Item Selection Criteria ................................................................. 232
Table A4  Item Final Refinement .................................................................... 233
Table A5  ASKp30 Rasch Model .................................................................... 237
Table A6  Complete Item History .................................................................. 238
Table A7  Summary Score Calculations .......................................................... 273
Table A8  Repeat of Table 10 (ASKp53 Rasch Model) In Item Entry Order ...... 274
LIST OF FIGURES

Figure 1  Methods Flow Chart ................................................................. 81
Figure 2  Distribution of Disability Scores ............................................... 85
Figure 3  Distribution of Response Option Frequencies .......................... 91
Figure 4  Distribution of ASK Scores by Severity of Disability ............... 99
Figure 5  Relationship Between the ASKp and CHAQ ............................. 100
Figure 6  Relationship Between Capability & Performance (Binary Responses) 109
Figure 7  Relationship Between Capability & Performance (Five Responses) 110
Figure 8  Overview of Trading-Off Process ........................................... 118
Figure 9  Law's Model of Childhood Disability ....................................... 162
Figure 10 Integration of Disability Models ............................................. 172
Figure 11 Conceptual Location of Paediatric Outcome Measures .............. 175

Appendix Figures

Figure A1 Response Patterns & Probability Predictions .......................... 215
Figure A2 Sample Item Response Curves ................................................. 224
Figure A3 ASK73 Development Flow Chart ............................................. 241
CHAPTER 1
INTRODUCTION

Health, the provision of care, and evaluations of the effectiveness of care have been the focus of much research. This attention to health is appropriate if we consider the pervasive impact it has on the lives of individuals who experience disorders of health. Children with musculoskeletal disorders are an example of a population in which there is relatively little documentation of the effectiveness of interventions. In addition, research has not adequately addressed all the consequences of disease. For example, there is a large body of literature on impairment (defined as a limitation in anatomic or physiologic function [World Health Organization 1980]), but much less has been written about disability (defined as a limitation in activity [WHO 1980]) and handicap (defined as a limitation in role function [WHO 1980]). Finally, while there is a movement towards client-centred approaches in health care, this has not been consistently reflected in evaluation methods that continue to use health records as a primary source of prevalence data. Therefore, the present research focused on a population, consequence of disease, and perspective seldom studied: disability among children as viewed through their own eyes.

During the 1950’s the polio epidemic, a life threatening and chronically disabling disease, drew attention to the plight of children with disabilities. In the late 1950’s the congenital deformities caused by Thalidomide again brought attention to children with disabilities. Agencies such as the Ontario Crippled Children’s Centre provided medical care for such children. Although the medical needs of children continue to make front page news on occasion (e.g., Siamese twins), the evaluation of the consequences of these disorders has received proportionately little attention.
Children are affected by a variety of disabling conditions including arthritis, congenital heart disease, cystic fibrosis, epilepsy, leukemia, and spina bifida. These conditions are often congenital and result in disability that remains with the individual for life. Thus, childhood disability is important because of its early onset and long duration. Childhood disability is also important because the limitations they cause are often not confined to the domain of the disorder responsible for the limitation (e.g., respiratory system effects). The effects often extend beyond the organ system and beyond the person to alter their ability to move about the community, be involved in social activities with their peers, and participate in a productive education. Thus, disability is not only a consequence of disease but is important because it is often a precursor to handicap [WHO 1980].

HISTORICAL BACKGROUND

Two dramatic changes in the demographic profile in Canada’s children over the last century have had a major influence on what we consider the important characteristics of health among children. The first of these changes was the reduction in infant mortality [White and Preston 1996]. This was largely due to the advent of antibiotics, improvement in nutrition, social circumstances and education. This change caused a decrease in the number of children dying from infectious diseases. However, part of the reduction in infant mortality occurred among babies born prematurely or with congenital anomalies (e.g., spina bifida). Unlike the neonatal survivors of infectious diseases who usually were left with few long term sequelae, the survivors of extreme prematurity and congenital anomalies more often experience long term impairments and disabilities. This increased incidence of long term sequelae in these latter two groups, coupled with increased life expectancy, caused a second demographic change: an increase in the prevalence of physical disabilities among children [Stanley and Blair 1994]. These changes resulted in an increased need to evaluate health related quality of life as opposed to mortality alone. The increased prevalence of chronic limitations among children resulted in a specific need to evaluate and plan for the associated disabilities and handicaps.
New methods were required to quantify childhood disability, handicap and quality of life. In contrast to mortality rates and life expectancy, that had been relatively simple to measure and track over time, indicators of health status are much less tangible. Most often those aspects of disease that technology had made it possible to measure (e.g., blood pressure, white count, and radiographic indices) were recorded, and then changes in health and quality of life were inferred from changes in these physiologic parameters. However, there was not a direct relationship between physiological (para-clinical) measures and health status [Huchison et al. 1996, Wilson and Cleary 1995]. Furthermore, the application of para-clinical measures did not adequately meet the needs of many groups (e.g., the military) and so subjective measures of health were sought [McDowell and Newell 1987]. The implementation of subjective measures was formidable, because health is a complex construct that is not easily measured. Many times measures were developed without using appropriate techniques or conceptual foundations.

Over time the field of health status has become an important area of scientific research. In the 1980's the World Health Organization introduced a classification of the consequences of disease that would act as a model of disablement [WHO 1980]. Inference from physiologic tests was no longer accepted as adequate [Viver et al. 1994]. The field of health care evaluation, termed Clinical Epidemiology [Feinstein 1982, Feinstein 1987, Feinstein et al. 1986], was growing. Some information to support this new science was borrowed from Epidemiology. Measurement principles were imported from the disciplines of education, where evaluation had a long history, and from psychology, where there had been a long tradition of measuring complex psychological traits [van der Linden 1994]. Health status became understood as the culmination of many components, including: mental health, physical function and social well being [WHO 1980].

From a review of measurement scales and historical accounts of influences on progress in disability evaluation [Albrecht 1992, McDowell and Newell 1996], it appears that the development and implementation of health status measures first focused on the most prevalent group of interest; adults. Political and economic influences were also at play, and these
influences also favored a similar group. Disability in the adult years constitutes a loss of productivity from a previously productive member of the labor force. Thus, adult disability is associated with a direct loss of produce for the employer and a loss of income coupled with treatment expenses for the individual. Government disability benefits and pensions reduced the direct adverse economic impact on the injured worker, however the broader economic impacts (cost to employers and health care systems) remained [Albrecht 1992, McDowell and Newell 1987]. These economic losses may have provided additional incentive to evaluate the effectiveness of health care interventions among adults. Progress in the evaluation of childhood mental health, physical function, and social well being has lagged behind progress in these fields regarding adults.

In the 1990's demographers, epidemiologists, politicians and health care providers began to react to the impact of assistive devices, the repercussions of new and expensive treatments (e.g., organ transplantation), and to prepare for the huge impact on our health care system from the aging of the baby-boomers (birth cohort of 1945 to 1960). Thus, health evaluation of adults, with added emphasis on the elderly, is likely to remain important. The disproportionately little research regarding the evaluation of childhood physical disability [Young and Wright 1995] underscores the importance of new research initiatives for this demographic group. To appreciate why childhood physical disability is worthy of study, it is important to understand the nature and magnitude of childhood disability in Canada.

STATEMENT OF THE PROBLEM

From a historical perspective relatively little attention has been directed towards health status evaluation in children. However, knowledge in this area is beginning to progress, particularly in Canada (e.g., [Boyle 1995, Cadman 1986, Gowland, 1991, Law 1988, Law 1992; Torrance 1995, Wright 1994a]). Further research is necessary to adequately evaluate the health status of children and the effectiveness of therapies for children with disabilities. The focus of this thesis is childhood physical disability.
Nature of Childhood Disability

"The Health of Canada's Children: A CICH Profile" [Hanvey et al. 1994a] from the Canadian Institute of Child Health depicts disability among children as being more often a consequence of congenital (or endogenous) rather than acquired (or exogenous) conditions, although both contribute [Hanvey et al. 1994b]. Although asthma, epilepsy and head injury are significant causes of disability among children, the present research is focused on children with musculoskeletal disorders. This population was selected in previous research that documented an absence of knowledge about the evaluation of physical disability in this particular clinical sub-group [Young 1994]. In this population, the most significant limitations are physical in origin (as opposed to mental or psycho-social [WHO 1980]). Thus, the focus of this thesis was restricted to the physical component of disability.

The most prevalent congenital conditions with a musculoskeletal component responsible for limitations in daily activities are cerebral palsy (2.3/1,000) and spina bifida (0.8/1,000) [Hanvey et al. 1994a]. Muscular dystrophy (primarily Duchenne MD) and haemophilia (that causes limitation due to bleeds into large joints), are present at birth but do not usually result in limitations in activity until at least age four, and contribute to a lesser degree. Osteogenesis imperfecta and Legg Perthes also contribute to disability in a small group of children. Absence of limbs also contributes (1.5/1,000), but the data do not distinguish between congenital versus acquired absence (e.g., traumatic loss). Furthermore, many children with these types of anatomic anomalies present only with impairments that may not translate into disability.

Among those with acquired musculoskeletal conditions causing disability are children with arthritis (1.1/1,000) [Hanvey et al. 1994a], acute injuries (e.g., fractures, dislocations), bone and soft-tissue tumors, and to a small degree children with scoliosis. Children with arthritis compose the largest group because of the longer duration of disability compared to the shorter duration of disability experienced by children with fractures and traumatic soft-tissue injuries.
Prevalence

The estimates of the prevalence of physical disability among Canadian children ranges from 4% to 7.8%. Based on the 1986 Ontario Child Health Study, Cadman et al. [1986a] estimated that 5.6% (confidence interval 4.2% to 7%) of children between four and 16 years of age experience limitations in activity due to physical impairments. In an analysis of the Ontario Health Survey 1990, Paul estimates that 7.8% of Ontario children between the ages of five and 15 suffer limitations in activity due to musculoskeletal disorders [Paul 1993]. The 1991 Health and Activities Limitation Survey (HALS) data indicated that 8.2% of Canada’s children from birth to 14 years of age have a disability that is cognitive, physical, emotional or sensory in nature. When only those who are “limited in normal activity” are considered, the revised estimate is 3.8% [Hanvey et al. 1994a]. Note that the confidence around the point estimates is wide because of the small proportion of children in these studies. American data from the National Health Interview Survey 1983-85 report that 6.2% of Americans five to 17 years of age experience limitations in activity due to chronic conditions [Albrecht 1992]. Thus, the estimate of the prevalence of musculoskeletal disability among five to 15 year old children lies somewhere between 4.0% and 7.8%, depending on the definition, survey, and questions used to estimate the prevalence of disability. Furthermore, because only one question was used in some cases, there may be a component of undetected physical disability in this population.

These childhood physical disability prevalence rates of approximately 6.5% suggest that childhood disability is less common than adult disability (approximately 22% prevalence of disability due to chronic musculoskeletal causes among Ontario residents 16 years of age or older [Badley et al. 1994]). Furthermore, children represent less than 20% of our population, and some musculoskeletal disorders of childhood abate during childhood (e.g., juvenile rheumatoid arthritis). This may partially explain the relatively small amount of research on measurement of childhood disability.
Although childhood disability is less common than adult disability, childhood disability is important for three reasons. (1) Many childhood disorders remain through adulthood (e.g., amputations and spina bifida) or become progressively more limiting over time (e.g., muscular dystrophy). Thus, the cumulative lifetime effects of childhood disability may far exceed those associated with disabilities acquired in the later decades of life. (2) Physical activity is an essential component of human development. Thus, childhood disability and inactivity may exert deleterious effects on other functional domains (e.g., mental and social). (3) The accuracy of prevalence estimates of childhood disability may be uncertain, since little research has been done on appropriate methods to measure physical disability in children [Gorter 1993, Young and Wright 1995]. Better measures may result in the detection of previously hidden disability, thus they may provide more accurate and higher estimates of prevalence.

Potential for Effective Treatment

The availability of treatment is also important in understanding the benefits of research into childhood disablement. There is a wide variety of effective treatment strategies available for many causes of childhood disability. For example: there are many surgical interventions to correct bony deformity, many medical interventions to control systemic lupus, and many rehabilitative techniques to promote recovery among trauma survivors. It is therefore imperative to determine exactly how effective these treatments are, and for which groups of children they are most effective. The determination of effectiveness is dependent upon a method for measuring meaningful outcomes, such as reduction in disability.

In many of these conditions this process of evaluating treatment effectiveness is made more difficult by very low prevalence rates and by the rapidly changing landscape of children’s functioning (due to growth and development). In addition, there is a clinical perception that these children have profound disabilities that must be treated immediately, not delayed until scientific evidence has had time to accumulate. However, while the age-related challenges and
the profound nature of childhood disabilities make evaluation complicated, they also enhance the potential gains.

Controlled clinical trials provide good quality evidence from which to evaluate the effectiveness of treatments. However, controlled clinical trials are rare in the areas of paediatric orthopaedics, rheumatology and physical therapy. Between January 1992 and May 1996 only 2.2%, 6% and 5.4% of paediatric articles published in the Journal of Bone & Joint Surgery (American), Archives of Physical Medicine & Rehabilitation, and Journal of Rheumatology respectively, were controlled clinical trials. In total, these numbers represent 20 articles over a five year period, with only three [Bednarczyk and Sanderson 1995, Edelstein and Berger 1993, Thornby and Krebs 1992] of the 20 including some form of functional outcome measure, and an additional three [Abud et al. 1993, Silverman et al. 1994, Wallace and Sherry 1992] including a clinician’s global rating. The numbers are significantly lower if only randomized controlled trials are included.

This deficiency of evidence of treatment effectiveness may be attributed in part to the lack of appropriate outcome measures [Young 1994]. While many developmental scales exist [Gowland et al. 1991], the purpose of treatment is not necessarily to restore these children to a ‘normal’ developmental trajectory, but to enable them to function better in their community. Since the intent of many interventions for this group is to minimize physical disability [Piper 1988], a method of assessing and monitoring physical disability is required.

In summary, there is a sizable population of children with musculoskeletal disabilities and many potentially viable treatment strategies. Thus, in addition to working on developing new treatment strategies, a collaborative approach to the problem of childhood disablement may aim to improve methods of evaluating existing treatments and to improve the conceptualization of childhood disability. Valid, reliable, responsive and feasible measures of childhood physical disability are essential to such an approach.
Potential to Evaluate Outcome

In a review of the childhood disablement literature I documented the absence of a valid, reliable, and responsive measure of physical disability for five to 15 year old children (see section on "Existing Measures" in Chapter 2 for more detail) [Young and Wright 1995]. Because of this dearth of appropriate methods for outcome evaluation, the evaluation of physical disability in children has been the focus of my research since 1992. A review of the literature was followed by the initial development of the Activities Scale for Kids (ASK) that was completed in 1994 [Young 1994, Young et al. 1995]. The ASK was designed to meet the needs of outcome assessment in future clinical trials and clinical practice. It was developed with careful attention to this purpose, and shows promise of fulfilling the outcome assessment needs of future clinical trials relating to paediatric musculoskeletal disorders. However, the ASK required additional testing to confirm its validity, determine its responsiveness, and refine the item content to enhance its feasibility.

The ASK was designed with particular attention to its conceptual basis [Young 1994]. At the time of the ASK’s development, a mechanism for exploring the conceptualization of childhood disability was foreseen as a beneficial adjunct. One of the conceptual issues of particular interest in disability measurement was whether scales should measure capability (what one can do) or performance (what one does do). Thus, the ASK was developed with both a capability and a performance version, so that it might be used to explore the implications of different measurement methods.

SUMMARY OF PURPOSE

The present research was a continuation of my previous childhood disability measurement research [Young 1994]. The purposes of the present thesis were: (A) to evaluate the measurement properties of the ASK (both at the summary score level and at the individual item level), and (B) to explore the magnitude and nature of the difference between what children
'can do' (termed capability) and what they 'usually do' (termed performance). The first purpose was achieved by assessment of item characteristics, further assessment of validity, and determination of responsiveness. The second purpose was achieved by statistical comparison of ASK-performance and ASK-capability scores, and through qualitative analysis of interviews with children regarding factors affecting their performance.

These primary objectives were inter-related, in that the first objective would enhance both the measurement properties of the ASK (thus enable better measurement of childhood disability) and provide the opportunity to measure the effect of contextual factors on disability estimates. The second objective would improve the understanding of the ASK, and improve the conceptualization of childhood disability by exploring factors that relate to the expression and detection of disability.
CHAPTER 2

REVIEW OF LITERATURE

The concepts of disability, measurement, and environmental factors that influence the expression of disability are central to this thesis. These concepts are reviewed in detail in this chapter. In addition, this chapter details the development of the Activities Scale for Kids (ASK), discusses its relationship to other paediatric measures, its conceptual basis, purpose, and accommodation for interactions with the environment.

CONCEPTUAL FRAMEWORK & DEFINITION OF DISABILITY

There are several conceptual frameworks that are pertinent to the present research. The most notable of these are the World Health Organization model (also known as the International Classification of Impairments, Disabilities, and Handicaps or ICIDH model) [WHO 1980], the Nagi model [Nagi 1965], and several variations or interpretations of these models. The most relevant of the variations are Verbrugge’s environmental model [Verbrugge and Jette 1994] and Kopec’s activity-space model [Kopec 1995].

Previous research suggested that the phrasing children use to articulate the consequences of musculoskeletal disorders was most consistent with the ICIDH model [Young 1994]. For example, they described their experiences of arthritis and Legg Perthes in terms of limitations in dressing, walking, playing, and other activities rather than in terms of not being able to bend their joints or having too little strength. Furthermore, children talk about the limitations occurring in a variety of community environments [Young 1994] including at home, school and in their neighborhoods. The effect of different environments on disability is an important aspect of the present research, and is also consistent with the ICIDH model. For these reasons, the ICIDH model was selected as the conceptual basis for the present research. The
ICIDH model is discussed in the following paragraphs. For the sake of comparison, a brief analysis of the Nagi [1965, 1991] models, and modifications to this model suggested by Verbrugge et al. [1989] and Kopec [1995] are also presented. More detailed descriptions of the Nagi, Verbrugge et al., and Kopec perspectives are included in Appendix I.

The ICIDH Model

The ICIDH model classifies the consequences of disease into three hierarchical categories: impairments (lowest level), disabilities (intermediate level), and handicaps [McDowell and Newell 1987, Townsend et al. 1990, WHO 1980, Wood 1980]. The model is consistent with the traditional “medical model” [Duckworth 1984, Swain et al. 1993], because of its strong relationship to diagnosis and disease and direct linkages to the ICD-9 codes (used in classifying individuals’ diagnoses for medical records). It defines impairments as restrictions of physiological or anatomic structure or function (e.g., inability to bend the elbow), disabilities as restrictions due to impairment of ability to perform activities within the range considered normal (e.g., inability to walk), and handicaps as disadvantage due to impairment or disability that limits the fulfillment of a usual role (e.g., inability to work or attend school) [WHO 1980].

In addition to the classification of impairments, disabilities and handicaps, the ICIDH framework has a second dimension of classification. On this second dimension, consequences of disease are classified according to: physical, mental, or psychosocial domains of health [WHO 1980]. Thus, the classification accommodates multi-faceted responses to disease, such as limited walking concurrent with social isolation that may occur in response to tumor diagnosis and treatment.

In comparison to the ICIDH model, Nagi’s model [Nagi 1965, Nagi 1991] has four levels: pathology (roughly equivalent to the ICD-9 codes), impairment (consistent with the ICIDH’s impairment level), functional limitations (roughly consistent with the ICIDH’s disability level but a little closer to impairment, e.g., inability to lift arm overhead) and disability (roughly consistent with the ICIDH’s handicap level, but a little closer to disability).
Important measure is appropriate because it will assess the immediate psychological response examples: When the purpose of measurement is to guide the administration of drug dosage, an example is all important in health care research, depending on the purpose of the research. For

From a conceptual basis, it can be seen that "imperative" disability and handicap measures

[Panel and Lewis, 1987]. Thus, new measurement approaches were sought.

is done, they are not appropriate to evaluate treatment effectiveness in clinical populations mortality and psychosocial morbidity may be useful for determining how the entire population morbidity and mortality rates. However, most psychiatric nurses believe that conditions are not mortality and morbidity rates. However, most psychiatric nurses believe that conditions are not groups. At the population level, effectiveness of interventions has often been inferred from statistical analyses, and is less reliable to conduct in a rigorously standardized manner for large assessments of individual cases. However, this approach is subject to bias, is not amenable to

Historically, the effectiveness of health care interventions has been inferred from clinical

Conceptual Basis of Therapeutic Effectiveness Evaluation

NICE or ICIDH models

much more interventional model of physical disability than previously presented by either the exception curve common in psycho-physical models. Thus, Hooke and Yerkesen envision a psychological factors. The presence of a different curve, that the lateral to the percipient activity is a function of the impairment level combined with the external circumstances and Hooke's Activity Space Model (Hooke 1995) suggests that the actual performance of an place. For example, concept of disability lower on the hierarchy than the ICIDH's handicap level, and inter-relatedness and feedback loops necessary between levels within the NICE model, and component should be incorporated into NICE's model. Verbrugge et al., go on to emphasize the depression activities that are considered with NICE's distinction between functional limitations, both Verbrugge et al. and Hooke distinguish between a level of situation-free and situation-
to the drug. Whereas, when the purpose is to evaluate program effectiveness, a disability measure may be more appropriate because program effectiveness requires an effect be observed at the level of the organism (or person) not just the organ. When the purpose is to determine community resource allocation, a handicap measure may be more appropriate since measures at this level incorporate societal impact. In some situations more than one type of measure may be required. For example, management of an individual’s inability to climb stairs (physical disability) may require the use of an impairment measure to determine aspects of the etiology and a disability measure to determine the impact of the limitation in the person’s usual environments. A handicap measure may also be indicated if the goal is to determine the impact of the limitation on the individual’s societal role (e.g., if the child had a job delivering newspapers the limitation may result in a significant handicap).

What is different about the purposes in these examples is the level at which the evaluation takes place, and thus the degree to which outside influences (i.e., effects from systems other than the musculoskeletal system or external to the person) are tolerated. Impairment assessments occur at the organ level and are useful when the purpose of the measurement is very specific (e.g., did the treatment affect bone density). Impairment measures derived by standardized techniques are affected less by other systems (e.g., mental function) or outside influences (e.g., bone density measured in a crowded room will not differ from that measured in another environment). However, because of the specificity to the organ (e.g., bone) the measures may not be generalizable beyond the organ [Huchison 1996], and therefore cannot inform us about the patient’s functioning in their community.

Disability assessments occur at the organism or person level and are useful when the purpose of the measurement is to determine whether the treatment altered the person’s functioning as a unified being. Such measures are usually developed to measure one domain of functioning in particular (e.g., physical function), but are not as exclusive to that domain as are impairment measures. For example, physical disability measures evaluate physical function primarily, but are affected to some degree by mental and social function (e.g., may be
influenced by cognitive abilities to overcome physical limitations). These other functions are an essential component of the person’s function in the community. Because of their inclusion, disability measures are generalizable to other similar environments, but may change if the environment is different in an important way (e.g., stair climbing disability may increase if the stairs are crowded).

In contrast, handicap assessments occur at the societal level and are useful when the purpose of the measurement is very broad. They incorporate the influences of physical, mental, and social factors and are global indicators of success (e.g., ability to work successfully in their desired occupation) [Badley 1987]. As a result, they may be more generalizable. However, handicap measures may measure more than health care interventions are responsible for changing. For example, a community’s outlook concerning amputees may have a great impact on handicap. The effect of the community will alter handicap, and may cause there to be an unexpectedly large increase in handicap level for a child with a lower extremity tumor after amputation and prosthetic fitting. The attitudes and prejudices of communities are only remotely affected by health care interventions. Thus, the influence of outside factors on handicap may introduce undesired noise into the measurement of a specific domain.

The use of a conceptual framework to determine the appropriate type of measure suggests impairment measures are often too specific (contain too little information) and handicap measures are often too broad (contain too much extraneous information or noise) to evaluate treatment effectiveness. Disability measures assess the functioning of the individual with the incorporation of some external influences, generate answers that are generalizable to the individual’s usual situation, and are consistent with the words children used to express the salient consequences of musculoskeletal disorders. Thus, disability measures were preferred for the purpose of evaluating therapeutic effectiveness.
methods and models may be used to describe populations, test hypotheses, and generate theory.

The purpose of measurement is to assign numeric values to attributes so that mathematical

METHODS OF SCALE DEVELOPMENT

tics conceptual foundation. Important that the methods by which distribution measures are constructed are consistent with the discipline level. In summary, the conceptual basis of health status has been introduced, and several

clinical needs is limited. Thus, this assessment further confirms the need for more research in measures (Haley et al., 1992; Russell et al., 1989), the integration of disability measures into the evaluation of treatment effectiveness. Thus is consistent with the general case of application and research practice was also considered. Implication measures have been well integrated into clinical practice. The degree to which each of the three types of measures have been integrated into clinical practice is limited.

The current is the most advanced level at which to conduct new research follow the same hierarchical model as do the ICIDH consequences of disability, with disability tracked in the hierarchy. Thus, the development of measurement knowledge is proposed to development of measurement techniques for disability is expected to facilitate learning about measurement of disability will facilitate our disability measurement research, and the measurement of disability. Because the ICIDH model is hierarchical, our knowledge of evaluation is to know and develop most efficiently, more needs to be learned about the curve, and our understanding of handicap in children is just beginning. If health care efforts over the next two decades, our understanding of childhood disability is at the base of the steep part of the current state of knowledge regarding impairment, disability, handicap. While information

The need for research into clinical outcomes was also considered with respect to the
[Nunnally and Bernstein 1994]. These abilities are essential for the generalizability of knowledge to samples beyond the one on which the evidence was gathered.

In many areas of medicine the development of measures has been aided by the readily observable nature of the constructs to be measured (e.g., DEXA scans to measure bone density). The necessity for a theory of scale development spawned from the need to evaluate characteristics of people that were not readily observable. It may be considered a circular argument: because the characteristic is not observable it requires a specially designed measure, however, such a measure cannot be validated unless it can be compared to something observable (e.g., histology of bone sections). Thus, a new body of conceptual knowledge was essential to meet the challenge of scale development when the attribute of interest was not fully visible. It is these methods that are applicable to the development of a measure of childhood physical disability.

There are two main strategies for the development of physical disability measures: clinimetric and psychometric methods. Both methods are based on a conceptual framework (e.g., the ICIDH). The concept is operationalized to a form that can be measured. In adult research physical function (the positive corollary of physical disability) has been operationalized as the “ability to perform self-care activities, mobility, and more strenuous physical activities” [Ware 1987]. In children, physical disability is demonstrated by a limitation in activities such as feeding, dressing, ambulation, and playing. Thus, childhood physical function has been operationalized as mobility, manual performance, leisure activities, instrumental ADL’s and self-care ADL’s [Meenan 1987]. In the development of the Activities Scale for Kids (ASK) I operationalized physical function as the ability to use the neurological and musculoskeletal systems to interact with the environment in a purposeful way [Young 1994]. Therefore, physical disability is a limitation in these abilities. This definition includes the common domains from adult conceptualizations, incorporates play, and is consistent with the definition of physical disability measured by the Pediatric Evaluation of Disability Inventory [Haley et al. 1992] (with which the ASK was intended to be consistent). It was this
operational definition of physical function that formed the basis of the initial ASK development and the ongoing development presented in this thesis.

**Clinimetric Methods**

The term *clinimetrics* [Feinstein 1982, Feinstein 1987] encompasses the methods for developing new measures of health based on the opinions of clinicians and patients. Similar methods have been described by Guyatt et al. [1986]. Items for clinimetric indices are usually elicited from clinicians who are experienced with the construct and patient population of interest, and supplemented with key items from other relevant measures [Davis et al. Submitted]. A recent review of the literature has documented a trend towards a more client-centred approach over the more clinician-centred approach [Kidd and Yoshida 1995].

The goal of clinimetric measure development is to ensure representation of all relevant aspects of the complex clinical concept. Thus, heterogeneity within the item pool is usually the primary focus, rather than homogeneity. The final items are selected and weighted according to the consensus of clinicians based on clinical relevance, although patient preferences are sometimes included (e.g., the Patient Specific Index [Wright et al. 1994b]). Summary scores are aggregated for groups of items that clinicians believe are related to one concept. The summary scores may be frequencies, the sum of weighted ratings, or on rare occasions are formed by Boolean logic (e.g., TNM staging system for the cancer staging) [Wright and Feinstein 1992].

Traditionally, clinimetric scales are created through comparisons to clinical judgment and evaluated for responsiveness to clinically important change. They were not consistently assessed for reliability or other forms of validity [Wright and Feinstein 1992]. A prime example of this is the Karnofsky scale, that was a multiple trait scale developed by one clinician based on his own experience [Karnofsky and Burchenal 1949]. Although it has questionable measurement properties [Hutchinson et al. 1979], it has been found to work clinically and therefore is used. More recently, however, reliability and validity assessment methods have
been applied to clinimetric indices (e.g., the Western Ontario McMaster Osteoarthritis Index [Bellamy et al. 1988]). In such cases, validity is often ascertained in comparison to clinical variables (e.g., joint counts, laboratory values or clinical judgment of disease severity), and reliability is often evaluated by a repeated measures design in stable subjects.

Psychometric Methods

The field of psychometrics has developed primarily within the field of psychology. Like clinimetrics, psychometrics encompasses a group of related methods for scale construction [Nunnally 1978, Nunnally and Bernstein 1994]. However, unlike clinimetrics the constructs of interest (states and traits) are often much less tangible (e.g., autonomy and intelligence rather than health status). Psychometrics also differs from clinimetrics in that it takes a much more structured statistical approach to development (e.g., the Guttman, Thurstone and Likert approaches) rather than the unstructured ‘dissected intuition’ [Feinstein 1987] approach of clinimetrics [Wright and Feinstein 1992].

Psychometricians derive items from a review of the literature and from existing item banks (i.e., from other measures), resulting in a large item pool [Davis et al. Submitted]. The intent of item reduction is to create homogeneous scales or sub-scales. Thus, statistical methods, such as factor analysis and scalogram analysis, are the primary methods used. Items are aggregated into a scale by a variety of statistical methods depending on the approach being employed (e.g., Guttman method). Similarly, weights (if used) are also based on statistical analyses. Reliability is often inferred from internal consistency (e.g., split-halves or Cronbach’s alpha) [Nunnally and Bernstein 1994], rather than by repeated measures designs. Validity is dependent upon homogeneity, thus scale structure (defined as the relationship between items) is ascertained by confirmatory factor analyses. Validity may also be assessed by convergence with conceptually related measures and divergence from unrelated measures [Cronbach and Meehl 1955].
In addition to clinimetrics and psychometrics, there is one additional aspect of scale development worthy of consideration: item characteristic analysis.

**Item Characteristic Analysis**

Although item characteristic analysis (ICA) is sometimes considered part of psychometrics [Nunnally and Bernstein 1994, Wright and Feinstein 1992], its origin was in education and its development was oriented towards a different objective (i.e., the assessment of the properties of individual items in relation to each other). Thus, it is considered separately here. ICA is based on item response theory [Hambleton et al. 1991, Molenaar 1995, Wright and Masters 1982]. ICA is not a method for the development of a measure from start to finish, nor a method of evaluating summary scale properties. Rather, ICA is a method for evaluating the behaviour of items and their associated response scales across a group of persons, and the behaviour of persons across a group of items. Thus, ICA analysis of items may be applied as an adjunct to either clinimetric or psychometric methods as a diagnostic tool in scale development.

ICA is based on two assumptions about the data: that the items are unidimensional, and that there is local independence among items after accounting for the dominant dimension (disability) [Hambleton 1991]. (Unidimensionality is defined as all items belonging to the same domain or concept, and local independence means that items do not share common characteristics other than the domain they measure.) ICA looks at the responses of individuals with different levels of difficulty across items of different levels of difficulty, and looks for unexpected response patterns (e.g., very able-bodied subjects reporting difficulty on very easy activities, or subjects with severe disabilities reporting no difficulty on very demanding activities). It is a probabilistic model, with the probability of a positive response taking on an ogive or sigmoid shape (similar to a cumulative normal distribution). The details of item response theory and Rasch modeling are provided in Appendix II.
ICA can be performed by several methods, all of which plot the probability of success on an item along the y-axis against the difficulty level of the item along the x-axis. The one-parameter model is called the Rasch model [Rasch 1960]. Under this model all items have the same slope and intercept but differ in difficulty [Hambleton 1991, Nunnally and Bernstein 1994]. The items' relative difficulty levels are expressed as the logit value (log odds units) at which their ogive curves cross the probability level of 0.5 in the case of binary items. In the case of ordinal response options, the difficulty level is expressed as the logit value at which the middle response option crosses the 0.5 probability level. The two-parameter model permits both the difficulty level and the slope of the probability ogive curve to vary between items, while the three-parameter models permits the difficulty level, slope, and the intercept to vary between items [Hambleton 1991, Nunnally and Bernstein 1994]. While there are some benefits to the two- and three-parameter models over the one-parameter model, from a practical point of view there are mathematical problems with the two- and three-parameter models that demand much larger sample sizes [Molenaar 1995] and result in frequent non-convergence [Linacre et al. 1996]. Thus, the one-parameter Rasch model, computed via the Bigsteps® computer program [Wright and Linacre 1992, Linacre and Wright 1993, Wright and Masters 1982, Wright and Stone 1979] was selected for use in this thesis.

ICA not only determines the difficulty level of items, and hence orders the items hierarchically, but it also determines the adequacy of fit of each item and each person to the model [Wright and Linacre 1994, Wright and Linacre 1989, Wright and Linacre 1992]. Fit statistics are subdivided into infit and outfit statistics, and the same criteria apply to both. Infit describes the variation of response patterns of an item for persons close to the item's difficulty level (e.g., unexpected responses on an item of above average difficulty by persons of above average ability). Outfit describes the variation of response patterns of an item for persons far from the item's difficulty level (e.g., unexpected responses on an item of above average difficulty by persons of below average ability).
The analyses of fit are important in item selection and refinement, since they indicate which items have excessive variability in response patterns (termed ‘noise’ and defined by a fit statistic >1.8) and items that are very similar to other items in the scale in their response patterns (termed ‘redundant’ and defined by a fit statistic <0.5) [Wright and Linacre 1994]. Because of the detailed analysis of item behaviour, Rasch modeling is useful in amending scales to improve their item content, fit to the model, and thereby reduces the amount of extraneous noise in the scale. In addition, when all items fit the model unidimensionality is confirmed [Fisher et al. 1995, Hambleton 1991, Linacre 1996, Nunnally and Bernstein 1994].

**Summary of Scale Development Methods**

Classic clinimetric measures are primarily expert opinion-driven measures of health. Classic psychometric measures are primarily statistics-driven measures of intangible traits. Both clinimetrics and psychometrics are focused primarily on evaluating the total score, while item characteristic analysis methods look at response patterns at the individual item level. Although these three approaches have different origins, theoretical bases, and methods, they all have a similar objective: to accurately and appropriately quantify a construct.

Clinimetrics and psychometrics are also similar in that both consider three phases of development: item generation, item reduction and summary score aggregation. While psychometrics consistently includes a fourth phase, testing of scale properties of validity and reliability, a similar stage is becoming more common in clinimetrics. Furthermore, a fifth stage of measure development, identification of scale structure, is shared by psychometrics (achieved by factor analysis) and item characteristic analysis (achieved by Rasch modeling). Thus, clinimetrics, psychometrics, and item characteristic analysis may be considered complimentary approaches [Davis et al. Submitted].

Recently, concepts from psychometrics have been integrated into what was traditionally the exclusive domain of clinimetrics: health care evaluation (e.g., the Sickness Impact Profile and SF36). Even more recently, both psychometric and item characteristic analysis methods
have found their way into paediatric health evaluation. For example, the Pediatric Evaluation of Disability Inventory [Haley et al. 1992] and the Functional Independence Measure [Granger et al. 1989] are both health status measures that have integrated psychometric and ICA methods in their development and evaluation. Consistent with recent trend, this thesis also draws on clinimetric, psychometric, and ICA methods in the evaluation of the ASK.

Regardless of the conceptual basis of a measure and the methods chosen for its development, the purpose of the measure and criteria for its evaluation must also be considered [McDowell and Newell 1987] in the scale development process.

**Purposes of Measurement**

The purposes of measurement have been described by several authors [Feinstein 1987, Kane and Kane 1981, Kirshner and Guyatt 1985, Williams and Naylor 1992, Wright and Feinstein 1992]. The definitions of Kane and Kane [1981] have been selected for the purposes of this thesis. They describe five purposes or major functions of measurement: *description, screening, assessment, monitoring, and prognosis*. These purposes of measurement were intended to be hierarchical as is shown in Table 1.

**Table 1** The Five Purposes of Measurement

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Level</th>
<th>Application</th>
<th>Essential Prerequisites (cumulative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prognosis</td>
<td>1</td>
<td>prediction of future status</td>
<td>predictive validity</td>
</tr>
<tr>
<td></td>
<td>(highest)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitoring</td>
<td>2</td>
<td>repeated assessments over time</td>
<td>test-retest reliability, fine degree of responsiveness</td>
</tr>
<tr>
<td>Assessment</td>
<td>3</td>
<td>determination of degree of a trait in sufficient detail to guide treatment</td>
<td>criterion validity, maximize sensitivity and specificity, measures the full scope with small increments</td>
</tr>
<tr>
<td>Screening</td>
<td>4</td>
<td>identification of those potentially at risk, and exclusion of those certainly not at risk</td>
<td>construct validity, minimal false-negative rate, distinguishes clearly between positives and negatives</td>
</tr>
<tr>
<td>Description</td>
<td>5</td>
<td>measurement of status</td>
<td>face validity, internal consistency (preferably construct validity)</td>
</tr>
<tr>
<td></td>
<td>(lowest)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Summarized based on Kane and Kane [1981]*
Table 1 also defines the terms used by Kane and Kane. As one ascends the hierarchy, the demands for documentation of measurement properties increase. Thus, a measure capable of prognosis is expected to meet the less demanding requirements necessary for description, but the converse may not always be true.

In addition to presenting the purposes of measures, the criteria for a good measure (e.g., construct validity) are also mentioned in Table 1. The following section discusses these criteria in detail, since they are an essential precursor to reviewing the literature on existing childhood disability measures. Furthermore, knowing the detailed criteria upon which a new measure will eventually be judged is one way to increase the likelihood that the new scale will eventually achieve success.

**Criteria for Grading the Quality of Measures**

The attributes of reliability and validity are the requisites of a good measure [Cole et al. 1994, Johnston et al. 1992, Kane and Kane 1981, McDowell and Newell 1987, Nunnally and Bernstein 1994, Streiner and Norman 1989]. Depending on the purpose of the measure, responsiveness [Streiner and Norman 1989] and feasibility [Feinstein 1987, Kane and Kane 1981] may also be important credentials, but are not in themselves sufficient conditions for measurement. Feinstein articulated three main areas of scale competency: *consistency* (hereafter referred to as *reliability*), statistical forms of *validity*, and *sensibility* [Feinstein 1987]. In this classification responsiveness may be subsumed under validity (i.e., where the construct is clinically important change), and feasibility is subsumed under sensibility.

For the sake of clarity, three main criteria are defined here. *Reliability* is defined as the degree of stability exhibited by a measure [Last et al. 1995]. *Validity* is defined as the degree to which a scale measures what it purports to measure [Last et al. 1995]. *Sensibility* is a term unique to Feinstein [1987]. It is defined as an "enlightened common sense" appraisal of a scale based on knowledge of pathophysiology and clinical reality [Feinstein 1987]. Sensibility
incorporates all the qualitative forms of evaluating a measure, that cannot be determined statistically. Feinstein's criteria for appraising clinical measures are shown in Table 2.

Table 2 Feinstein’s Criteria for Appraising Measures

<table>
<thead>
<tr>
<th>Main Criteria</th>
<th>Sub-Criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity (statistical)</td>
<td>1 Predictive Validity</td>
<td>Strength of association with a future event or measure</td>
</tr>
<tr>
<td></td>
<td>2 Criterion</td>
<td>Comparison to an accepted standard</td>
</tr>
<tr>
<td></td>
<td>3 Construct</td>
<td>Relationship to an accepted theory</td>
</tr>
<tr>
<td>Reliability (statistical)</td>
<td>4 Inter-Rater Reliability</td>
<td>Kappa values and intra-class correlation coefficients between different raters</td>
</tr>
<tr>
<td></td>
<td>5 Test-Retest Reliability</td>
<td>Intra-class correlation coefficients from a repeated measures design</td>
</tr>
<tr>
<td></td>
<td>6 Internal Consistency</td>
<td>Cronbach’s alpha</td>
</tr>
<tr>
<td>Sensibility (common sense)</td>
<td>7 Purpose &amp; Framework</td>
<td>Appropriateness</td>
</tr>
<tr>
<td></td>
<td>8 Content Validity</td>
<td>Comprehensiveness, weighting and suitability of summary scores</td>
</tr>
<tr>
<td></td>
<td>9 Overt Format</td>
<td>Comprehensibility, replicability and suitability of rating scales</td>
</tr>
<tr>
<td></td>
<td>10 Face Validity</td>
<td>Coherence and suitability of the components</td>
</tr>
<tr>
<td></td>
<td>11 Ease Of Usage</td>
<td>Time, cost and expertise required to apply</td>
</tr>
</tbody>
</table>

Summarized based on Feinstein [1987a]
Note: the numbers beside the sub-criteria are carried over into Table 3.

Relationship Between Purpose of Measure and Criteria for Evaluation

When the purposes of measurement (Table 1) are compared to the criteria for evaluation (Table 2), it is apparent that there is not a one to one relationship of criteria to purposes. However, some hierarchical progression can be seen in both Kane and Kane’s purposes and Feinstein’s criteria for scale quality. For example, sensibility is a basic criteria that may be
applied without the necessity for data collection, thus can be easily evaluated and should be met by a scale used even for descriptive purposes (lowest level on the purpose hierarchy). Furthermore, some aspects of sensibility are generic and generalizable across multiple purposes (e.g., comprehensibility), thus may have been previously determined. In contrast, reliability ranks somewhat higher, requires more complex evaluation methods, and is specific to a given purpose. Thus, while sensibility alone (including face and content validity) may be sufficient for some descriptive purposes, the statistical forms of validity become increasingly important as you ascend the purpose hierarchy. Responsiveness and test-retest reliability may rank higher still, and correspondingly they become essential at the monitoring level. All criteria, including predictive validity, should be met at the prognostic level.

A measure that meets all the criteria in Table 2 would be considered a good quality measure for at least one specific purpose and population. The additional documentation of responsiveness (measured by effect size [Cohen 1988], or standardized response mean [Fleiss 1986, Katz et al. 1992], etc.), becomes a valued quality at the level of monitoring.

Other Measurement Issues

Consideration of the purposes and quality of measures must also take into account the various sources of error. Error is introduced in two forms: random error and systematic error.

Random error is particularly problematic in the measurement of individuals as opposed to groups [McHorney and Tarlov 1995]. Measures must be more precise (i.e., reliability≥0.9 [Nunnally and Bernstein 1994]) and accurate (i.e., correlation with similar measures >0.8) if they are to be used to guide the management of individual patients, particularly when used to guide clinical decision making [McHorney and Tarlov 1995]. This is because random errors within one subject will not be offset by other random errors in the opposite direction in another subject. In contrast, slightly lower levels of these parameters are accepted for use in groups of subjects, because by definition random errors will have a mean of zero. Thus, increases in sample size can be used to compensate for random error in groups introduced by lower levels
of reliability and validity [Nunnally 1978]. Therefore, although the five purposes of measurement apply equally well to individuals and groups, the prerequisites for acceptability are generally more stringent for use with individuals than with groups.

Systematic error is of concern both in the measurement of individuals and groups. One common source of systematic error is respondent bias. For example, the perspectives of clinicians may be discordant with those of children with disabilities and their families, and thereby introduce a systematic error. Therefore, the purpose of the measurement must consider whose perspective is of greatest interest. Is the purpose to determine disability as seen through the eyes of clinicians or children? Previous research has suggested that the difference between child and parent-report has minimal effect in some settings [Young et al. 1995] and for some constructs [Doherty et al. 1993, Duffy et al. 1993], therefore negating the need to chose one perspective exclusively in the application of a measure. However, it is more often the case that little is known about the influence of perspective. Thus, the method of measurement must reflect the intended perspective as part of the purpose, to prevent systematic bias.

Systematic error may also occur due to missing data and the inclusion of not-applicable response options. True missing values occur when items are left unanswered. However, not-applicable responses are included in some questionnaires to permit the inclusion of some items that may not be relevant to all respondents. Measures that include not-applicable responses generally exclude both missing values and not-applicable responses from the numerator and denominator in the summary score calculation. If the frequency of missing or not-applicable items is related to severity (i.e., respondents choose either to not answer difficult questions or report them as not-applicable) a new source of systematic error may be introduced into the summary scores. However, if not-applicable responses are not available, there may be an equally concerning systematic error introduced by forcing patients to answer questions about activities they have never had the opportunity to try or by imputing values for these items. Thus, the presence or absence of not-applicable response options and mechanisms for dealing with missing values should be considered in any critical review of an outcome measure. The
appropriate mechanisms may vary depending on the intended use of the measure and its conceptual foundation (i.e., is it intended to be specific to the individuals environment or consistent across all respondents?) [McDowell and Jenkinson 1996].

Lastly, in measures intended to monitor change, error introduced between the first and second time point must be assessed. Since the detection of a significant change is determined by the magnitude of the change relative to the error, test-retest reliability is used to quantify the amount of true variability in subjects scores relative to the amount of error between the first and second time points. Minimization of error over time is crucial in monitoring measures.

In summary, the components to be considered in either scale selection or scale development include identification of: a population of interest (i.e., five to 15 year olds with musculoskeletal disorders), a conceptual basis for the measure (i.e., disability), a theory of scale development (i.e., combination of clinimetric, psychometric and item characteristic analysis), a purpose (i.e., monitoring), the quality of the measure (i.e., validity), the perspective of interest (i.e., child's own point of view), and consider the sources of error inherent in the measure.

Although Feinstein and Kane developed their criteria with adult and geriatric populations in mind, the criteria are applicable to paediatric scales and to all consequences of disease from impairment to handicap. The continued acceptance of impairment measures, and the consistent demand for evidence of quality before disability measures are applied, suggest that different standards may have been applied to these two types of measures. The reliability of disability measures may be as good or better than impairment measures if they were compared using the same yardstick. Furthermore, the sensibility and validity of disability measures may be superior to that of impairment measures, because disability measures may be more concordant with the purpose of the outcome assessments of health care programs. Since many scales exist for measuring physical disability in children, these are reviewed with respect to the criteria just discussed.
EXISTING MEASURES

The potential to use or adapt another existing measure was examined before embarking on the development of the ASK.

Relevant Measurement Approaches

The existing literature on both paediatric, adult and geriatric disability suggests that the existing scales may be classified according to their conceptual approach to measurement as generic, disease-specific, or domain-specific [Bowling 1995]. Generic measures are those that are not specific to a particular diagnostic group, and therefore are useful in comparing the degree of disability between different groups (e.g., rheumatoid arthritis vs. asthma). Disease-specific (or condition-specific) scales are designed to measure only those aspects of disability that are important or likely to change in one particular diagnostic group, and therefore are expected to be more sensitive to change in disability due to treatment for the underlying diagnosis. Domain-specific scales are designed to measure only one aspect of disability (e.g., physical disability) as opposed to all domains (e.g., physical, mental and social aspects). A scale may have more than one classification (e.g., may be a disease-specific measure of physical function, or a generic measure of health that incorporates all domains).

Within each of these classifications, measures may also be based on a either a standard set of items, or be individualized to those items that are relevant to the patient. The latter have sometimes been referred to as "patient-specific" measures. Scales with standard item sets include the Sickness Impact Profile [Bergner et al. 1981], Short Form 36 [Ware and Sherbourne 1992] and Functional Independence Measure [Uniform Data System 1987]. These scales generate a summary score based on all items and imputation must be used, or cases dropped from the analysis, if missing data account for more than 5% of responses. One of the advantages of standard item scales is that they may be used to generate population norms. One of the disadvantages is that they may contain items of little relevance to individual patients. Patient-specific scales, such as the Patient-Specific Index [Wright et al. 1994b], the MACTAR
[Tugwell et al. 1987], the Childhood Health Assessment Questionnaire [Singh et al. 1994] and the Canadian Occupational Performance Measure [Law et al. 1990], have the advantage of being more closely aligned with the issues of importance to the patients to whom the scale applies. Patients rate only those items that are relevant to themselves, and the remaining items are excluded from the summary scores. This results in variability in the number and content of items answered in patient-specific measures, and makes normative comparisons more complex. Furthermore, no criteria have been specified for the proportion of missing or not-applicable response options tolerated. Thus, either standard or patient-specific item sets may be used, but the advantages and limitations of each must be recognized.

Having provided an overview of the ways in which measures may approach the measurement of various characteristics of health, it was now worth examining the particular domain of interest and the issues unique to that domain in further detail. The present research was interested in physical function in particular. Measures of physical function may be further classified according to the specific pragmatic approach to the measurement of physical disability that was taken. There were at least two axes of classification: by method of data collection (observation vs. self-report or proxy report), and by type of information (capability vs. performance) [Young and Wright 1995]. Each of these pragmatic approaches had certain implications.

Observational measures (often referred to as direct or objective) may be most familiar and therefore appealing to clinicians, but are time consuming [McDowell and Newell 1987], expensive and have most often been conducted in the clinic setting rather than in the individual’s most important environments. Report measures (often referred to as indirect or subjective) may take into consideration the individual’s home and work environments, but are potentially subject to respondent bias [McDowell and Newell 1987].

Capability measures provide information on whether or not an activity could be completed [Haley et al. 1991a, Young et al. 1996] and may be sensitive to subtle changes if a graded
response scale is used. However, capability measures may miss out on information regarding the acceptability of that level of ability with respect to the everyday situations of the individual. Capability measures may also have the advantage of incorporating a measure of effort required [Feinstein 1986] or degree of difficulty. They may therefore be considered a measure of quality of physical function. Performance measures determine the actual physical enactment of activities, and therefore may collect this information with respect to the individual’s usual environments. However, performance measures may miss information on the degree of difficulty individuals experience performing at their reported level. When performance is graded in terms of frequency of activities, it may be considered a measure of quantity of physical function.

Finally, one alternative for the measure of either capability or performance was identified; that being the measure of disability in terms of independence. Such measures record the degree of assistance required to complete activities, and may indicate the resources required by that individual. However, the expectation of independence is not consistent with the concept of disability (e.g., as defined by the World Health Organization [1980]) for all age groups (e.g., not necessarily a reasonable expectation for small children and the elderly) [Haley et al. 1991a].

**Relevant Scales**

After a review of existing physical disability scales of all classification types [Young and Wright 1995], those most relevant to children with musculoskeletal disorders were considered to be: the Childhood Health Assessment Questionnaire (CHAQ), the Gross Motor Function Measure (GMFM), the Health Utilities Index Mark III (HUI3), the Juvenile Arthritis Functional Assessment Report (JAFAR), the Pediatric Evaluation of Disability Inventory (PEDI), and the paediatric version of the Functional Independence Measure (Wee FIM). Table 3 shows a critique of these scales, and the recently developed ASK, in terms of how they fit into Kane’s framework and Feinstein’s criteria (note: the numbers listed below the headings validity, reliability and sensibility correspond to those in Table 2 on page 25). Table 3
considers the scales with respect to their intended target population, not the population of five to 15 year old children with musculoskeletal disorders.

Table 3 Review of Existing Scales

<table>
<thead>
<tr>
<th>Measure</th>
<th>Target Population</th>
<th>Kane’s Purposes</th>
<th>Feinstein’s Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Validity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 2 3</td>
</tr>
<tr>
<td>ASK</td>
<td>5 - 15 years of age musculoskeletal</td>
<td>monitoring</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>CHAQ</td>
<td>1 - 19 years of age arthritis</td>
<td>monitoring</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>JAFAR</td>
<td>7 - 18 years of age arthritis</td>
<td>monitoring</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PEDI</td>
<td>0.5 - 7 years of age developmental delay</td>
<td>monitoring</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>GMFM</td>
<td>0.5 - 7 years of age cerebral palsy</td>
<td>monitoring</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>WeeFIM</td>
<td>0.5 - 7 years of age not disease-specific</td>
<td>assessment</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>HUI3</td>
<td>6 - 18 years of age general population</td>
<td>description (i.e. health surveys)</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
</tbody>
</table>

**LEGEND**

Validity: 1=predictive, 2=criterion, 3=construct
Reliability: 4=inter-rater, 5=intra-rater, 6=internal consistency
Sensibility: ✓ indicates this has proven satisfactory, X indicates this has proven unsatisfactory, ○ indicates insufficient data to evaluate, * documented in this thesis

It is also important to assess these scales relative to the specific population and purpose of interest in this thesis. A scale was sought that was capable of monitoring physical disability in five to 15 year old children with musculoskeletal disorders.

The CHAQ [Singh et al. 1994, Singh et al. 1990] is a 29-item measure of functional status among children one to 19 years of age with juvenile rheumatoid arthritis. It was developed from the adult Stanford Health Assessment Questionnaire (HAQ) [Fries 1991, Spitz and Fries...
1987] disability domain to evaluate the effectiveness of treatments. Thus, its purpose is monitoring of small groups. It is intended to be completed by parents or children (if greater than eight years of age), thus measures disability from the perspective of families. It has been proven valid with respect to Steinbrocker functional class (Kendall's tau = 0.77), number of involved joints (Kendall's tau = 0.67), morning stiffness ratings (Kendall's tau = 0.54) and Disability Index scores (rho = 0.84 [Singh et al. 1994]). It has subsequently been tested on children with dermatomyositis, and found to be reliable (ICC=0.87) and relatively responsive (responsiveness statistic 0.90) [Feldman et al. 1995]. However, because the items are specific to arthritis, its applicability as an indicator of prevalence and severity of musculoskeletal disability is limited. Specifically, it may underestimate disability in children with musculoskeletal disorders if their only limitations occur in activities of little importance among children with arthritis.

When the CHAQ was considered for use in the target population (five to 15 year olds with musculoskeletal disorders) several limitations were identified. It is specific to arthritis and was derived from an adult scale. Furthermore, although the HAQ was based on a conceptual model of disease, when the transition to a paediatric version was made the conceptual basis was not reviewed and documented. Finally, the sensibility of the CHAQ is compromised by: the aggregation methods (the summary score is the average of the worst score in each of eight domains, thus, information on less severe items is dropped from the scale as are missing values and not-applicable items), assignment of penalty points for using assistive devices, and potential for confounding of age and disability effects in the use of not-applicable response options.

The JAFAR [Howe et al. 1991] is a child-report or parent-report measure of physical disability for seven to 18 year old children with arthritis. It contains 23 tasks that were considered easily tested, basic to daily function and mastered in healthy children of the same age group. Each task is rated by the respondent in terms of frequency of performance on a three level ordinal scale. A visual analog pain scale is also included. The JAFAR was intended
to evaluate the effectiveness of treatments in this clinical sub-group, and to be an adjunct to its predecessor, the Juvenile Arthritis Functional Assessment Scale (JAFAS: an observational assessment of physical disability) [Lovell et al. 1989]. Thus, the JAFAR's purpose is to *monitor* physical disability from the perspective of families.

Unlike the other scales described, the JAFAR's weakness is derived from a number of small limitations rather than one primary limitation. It is disease-specific, and needs more rigorous testing of validity. It is restricted to basic competence level activities and tasks, and thus lacks items to assess slight disability experienced by children in activities such as running. It is aggregated as the sum of answers to all questions (range zero to 46 where 46=maximal disability), and is not explicit about how missing values are dealt with [Howe et al. 1991]. Although it appears to fit the ICIDH framework, the developers have not documented its intended conceptual framework. Thus, while it was unable to meet the objectives of monitoring physical disability in the target group of interest, it contained some valuable concepts. For example: the response options all of the time, some of the time, etc.

The PEDI [Feldman et al. 1990, Haley et al. 1991b, Haley et al. 1992] is described as a physical disability measure for children six months to seven years of age, but it is actually a series of measures for children with developmental delay and neurological conditions. It was developed to evaluate the effectiveness of treatments in this clinical sub-group. Thus, its purpose is to *monitor* physical impairment and disability in small groups from the perspective of clinicians, teachers and families. The measures within the PEDI are divided into a “Functional Skills” scale (described by the authors as consistent with *functional limitations* in Nagi's model) containing 197 items, a “Caregiver Assistance” scale (described by the authors as consistent with *disability* in Nagi's model) containing 20 items, and a “Modifications” scale (described by the authors as consistent with *disability* in Nagi's model) containing 20 items. The items in each of these scales were spread over the same three sub-domains (self-care, mobility and social function). The “raw” summary scores are calculated by summing items within each domain of the functional skills and caregiver assistance scales, yielding six
summary scores ("scaled" scores may also be computed based on a normative sample Rasch model). Scores for the modifications scales are reported as three frequency distributions. Missing data are not tolerated according to the administration manual [Haley et al. 1992].

The most significant limitation of the PEDI is its focus on the developmentally delayed population. Because it is disease-specific, it cannot be generalized to the broad musculoskeletal disorders of interest in this thesis. For example, it contains few higher level functional activities, frequently the only evidence of limitations in children with L5 spina bifida. However, the PEDI may be useful for children under seven years of age who have musculoskeletal disorders. (Reid et. al provide a detailed critique of the PEDI with respect to children under seven years of age [Reid et al. 1993]). It also appears that, although it was developed with very careful attention to the conceptual framework and measurement properties, the scales contains a mix of impairment and disability items.

The GMFM [Russell et al. 1989, Russell et al. 1990] is an 88-item measure of gross motor function for children with cerebral palsy. It is an observation measure of motor skills. In terms of the ICIDH guidelines, it covers some aspects of impairment and some aspects of disability. It was developed to evaluate the effectiveness of treatments in this clinical sub-group, with clinicians being responsible for the selection of the items, and the evaluation of children. Thus, its purpose is to monitor physical impairment/disability from the perspective of clinicians. It was developed through a systematic approach, taking into account the purpose of the measure and conceptual issues related to the evaluation of motor function from a clinical perspective. A single summary score is calculated as the average of eight sub-domain scores (sub-domain score = average of applicable item scores). Thus, although a not-applicable response option is not included, missing items are simply dropped from the score. Testing showed moderate relationships with constructs, excellent test-retest reliability, and good responsiveness.
The GMFM paid great attention to detail during its development, as is evident in the associated scoring methods and manual. However, when it was reviewed with regards to the target population for the present research, it was clear that the generalizability from cerebral palsy to a generic group of musculoskeletal disorders was not possible. In addition, it was found to be conceptually closer to the impairment level than the disability level, and it has limited feasibility because it must be completed by a trained observer.

The WeeFIM [Granger et al. 1989, McCabe and Granger 1990, Research Foundation-State University of New York 1991] is stated to be a measure of disability, in an unspecified clinical population. The target age group is six months to seven years. It was developed as a uniform data set to be gathered in all rehabilitation centres. Thus, its intended purpose is the assessment of large groups from the perspective of clinicians. It contains 18 items, each scored by a clinician during an interview with a caregiver. The summary score is the sum of all item scores and has a range of 18 (maximal disability) to 126, with no missing values tolerated.

The WeeFIM has many limitations. The greatest limitations are conceptual: although it purports to measure disability, its response options measure dependence. In their intended age group dependence is expected, thus should not be equated with disability. Also, penalty points are assigned for supervision. Thus, there is likely a significant age gradient, since supervision is usually provided for toddlers. Furthermore, the assignment of penalty points for the use of assistive devices may be conceptually inappropriate since disability is about activity not mechanics of movement, rendering the scale weak in several areas of sensibility (framework, face validity, and construct validity). This scale also has a preponderance of double and triple barreled questions and is costly to apply. Furthermore, although it does identify areas of dependency, this not in sufficient detail to suggest treatment strategies or detect small changes, particularly in children with mild and moderate disabilities.

The HUI3 [Boyle et al. 1995, Feeny et al. 1995, Patrick and Erickson 1993] is a measure of physical disability in children in the general population. This measure began as the HUI1,
that was developed to evaluate outcomes of very low birth weight babies using a multi-attribute framework [Cadman et al. 1986b]. The HUI1 contained six domains and generated 960 unique states, each with a utility value. The HUI2 was a seven domain modification, that was intended to assess health status in childhood cancer survivors [Feeny et al. 1992, Feeny et al. 1993]. It described 24,000 health states. The HUI3 was the most recent modification developed to collect demographic information on childhood disability in the 1990 Ontario Health Survey, and has also been used in the 1991 Statistics Canada General Social Survey, the National Population Health Survey and the National Longitudinal Health Survey of Children in Canada. It includes eight domains, and each domain is scored based on a series of Guttman-type questions (increasing level of disability across questions in each domain). The single scores from each in combination describe 972,000 unique health states. Thus, its purpose is to describe the multi-dimensional health status of the childhood population, from the perspective of demographers.

The limitations of the HUI3 with respect to monitoring physical disability are primarily related to the difference in purpose. Thus, the level of discrimination within the ambulation and dexterity domains are insufficient to meet these objectives.

Having recognized many of the deficiencies of existing scales it was apparent that a new measure was required. It was also apparent that many of the existing scales had some useful items and concepts. These were carefully selected for integration into the new measure.

OVERVIEW OF THE ASK

The Activities Scale for Kids (ASK) is a self-report measure. The justification for the development of the ASK is apparent from the preceding pages: no generic measure was available for children five to 15 years of age, that met minimal levels of sensibility. Although clinicians could globally define the extent of physical activity limitations (mild, moderate and severe), changes due to time or treatment were not likely to be detectable using such a coarse scale. Furthermore, none of the existing measures had fully integrated the perspectives of
children into the scale development process [Young 1994]. From a conceptual standpoint this latter limitation was serious, and precluded the adaptation of existing measures since the foundations of these were potentially flawed.

**Purpose of the ASK**

The purpose [Kane and Kane 1981] of the ASK was both to assess the status of childhood disability and to monitor change associated with time and treatment interventions among five to 15 year old children. Although some authors may argue that a scale cannot meet more than one purpose [Guyatt et al. 1986], the model developed by Kane and Kane suggests a monitoring measure will meet the criteria for assessment [Kane and Kane 1981]. (This assertion will be indirectly tested in the methods of this thesis.) The ASK was also intended to explore conceptual measurement issues, and hence had two versions. The ASK-performance was intended to measure what children ‘did do’, and the ASK-capability was intended to measure what children ‘could do’. The comparisons of data from the two versions would enable the exploration of the effects of two potentially different environments on disability measures.

**Clinical Population of Interest**

At the time of the ASK’s inception, the dearth of physical disability outcome measures was most obvious in the five to 15 year age range. In addition, this age range was attractive because the impact of growth and development was less of a concern in children over five years of age, and because it included some overlap with the PEDI. This overlap was desired to enable continuity across the transition from the PEDI to the ASK in the future. This would facilitate research crossing the five year of age lower limit of the ASK, and thus facilitate the ASK’s use in long term studies [Young 1994].

In the present research the definition of musculoskeletal disorders was very broad, including children with arthrogryposis, arthritis, Blount’s, congenital amputations, diplegic cerebral palsy, femur fractures, Guillain Barré, leg length discrepancies, muscular dystrophy, osteogenesis imperfecta, radius fractures, scoliosis, spina bifida, and tendon repairs. These
diagnostic groups represent most of the major causes of musculoskeletal disability in children. Thus, the magnitude of disability ranged from very mild to very severe, and sensitivity to change was important at all levels of disability.

**Conceptual Basis of the ASK**

The World Health Organization’s International Classification of Impairment, Disability and Handicap (ICIDH) [WHO 1980] formed the framework that supports the conceptualization of disability used in the ASK. Within this framework, the ASK was intended to measure the domain most consistent with the words used by children to express their health-related problems and also be consistent with the intended benefits of various treatments. Children present for treatment either complaining of ‘hurt’ or that ‘it’s hard’ for them to do their usual activities. Alternatively they present for treatment because an adult (i.e., mother) has detected an abnormality that they are worried might lead to future problems. Thus, the nature of their concerns lie both within the scope of impairment (pain and anatomical abnormality) and disability (activity restrictions) [WHO 1980] of predominantly physical origin. The primary intent of most treatments is the relief of these complaints. While physical impairment can be assessed accurately [Beattie et al. 1990, Cahalan et al. 1989, Gogia et al. 1987, Goldsmith et al. 1992, King-Thomas and Hacker 1987, Savedra et al. 1989, Wilkie et al. 1990], children’s physical disability cannot be accurately assessed for many diagnostic groups because of an absence of appropriate measures [Young and Wright 1995]. Thus, physical disability was selected as the focus of the ASK [Young 1994].

It is acknowledged that children with musculoskeletal disorders may also present for treatment because of handicap, and that there is a valid argument for the development of a handicap measure. This was not pursued for four reasons. (1) Most therapeutic interventions do not directly address handicap. Clinicians are, therefore, primarily interested in treating the impairment for the purpose of reducing disability, with the hope that this will translate into a long term reduction in handicap. (2) Short term and intermediate outcome assessments do not
leave adequate time for therapeutically induced impairment reductions to translate into handicap reduction. Moreover, long term assessments in children often become confounded with developmental influences. (3) There are pragmatic concerns at the present time with the operationalization of the WHO definition of handicap [Badley 1987] and overlap with disability. These have not yet been adequately resolved in the continuing evolution of the WHO framework. (4) Handicap is impacted upon by many extraneous factors (e.g., mental capacity) [Badley 1987], thus its use in clinical trials may introduce an unnecessary degree of variability into the outcome. Therefore, handicap was not the focus of the ASK. However, future research may ultimately lead to measures of how disability impacts on role performance and handicap measures may follow in time.

Since the objective of therapy is to improve the functioning of children in their usual community settings, the ASK was conceptualized as a measure of physical disability experienced in children’s usual communities. Furthermore, the ASK conceptualized physical disability as seen through the eyes of children rather than clinicians’ or parents’ eyes. Justification for these perspectives has been previously established [Young 1994, Young et al. 1995, Young et al. 1996].

**ASK’s Development to Date**

In the development of the ASK, clinimetric, psychometric, and item characteristic analysis methods were not looked upon as conflicting strategies, but rather as unique strategies, each of value, and potentially complementary in the process of scale development. Clinimetric strategies are most efficient for generating items for measures of health, thus were employed to generate items for the ASK. Clinicians’ judgments were used to complete the initial item reduction. This judgment occurred in a focus group and drew on their experience combined with data from children and parents on the frequency with which limitations were reported and their importance.
These initial steps culminated in a 73 item *developmental edition* of a childhood disability questionnaire 1994 [Young 1994, Young et al. 1995, Young et al. 1996] titled the *Activities Scale for Kids* or ASK. The items contained in this version of the questionnaire are shown in Table A6 of Appendix III. As previously mentioned, the ASK had two versions: performance and capability. Both were administered to 28 children, 18 of whom completed them twice within a two week period. Summary scores for each version were calculated as the sum of all items divided by the maximum possible score (range zero to 100 percent, where 100=full function). These summary scores permitted assessment of the ASK’s reliability and a preliminary assessment of validity. In summary, it was found to be reliable over repeat administrations (random effects intra-class correlation coefficient > 0.96), valid in comparison to clinicians’ global ratings of mild, moderate and severe (ANOVA p=0.002), and child-report was found to be concordant with parent-report [Young 1994, Young et al. 1995]. (An overview of the initial development of the ASK and testing is shown in Appendix IV.)

As can be seen in Table 3, even at this early stage, the ASK compares favorably to other childhood disability measures when critiqued according to the purposes of Kane and Kane and the criteria of Feinstein (sensibility, validity and reliability). It was also decided that because child-report and parent-report were concordant, children should be encouraged to answer the ASK independently, although parents are permitted to read questions to children under age 10. Based on these encouraging results, it was concluded that the ASK showed sufficient evidence of quality to warrant further development.

Although the ASK showed promise of fulfilling the outcome assessment needs of future clinical research relating to paediatric musculoskeletal disorders, it’s responsiveness was unknown and it’s feasibility limited (completion time of 30 to 60 minutes per version [Young et al. 1996]. Therefore the ASK requires additional testing and refinement. This testing was completed as a major component of this thesis, and is outlined in Chapter 3 (Methods).
ROLE OF THE ENVIRONMENT IN DISABILITY EVALUATION

One of the unique attributes of the ASK is its ability to assess the impact of different environments or contexts on physical function scores. The impact of environmental factors on disability has been an important aspect of the development and testing of the ASK since its inception. Because of a desire to explore contextual effects on disability, the ASK was designed to measure disability in two distinct (although not necessarily mutually exclusive) environmental contexts. The ASKc was designed to measure what one 'could do' and implied a hypothetical (and perhaps ideal) environment, whereas the ASKp was designed to measure what one 'did do' and explicitly stated that this was in one's usual (or real) environment [Young et al. 1996]. Initial comparisons of the two versions of the ASK in a small sample (n=28) showed that scores based on capability exceeded those based on performance (p<0.006) [Young 1994]. The discrepancy between ASKc and ASKp scores was attributed to a difference in context [Young 1994].

Some information on the role of context on the ASK was provided by the work of Sen in his discussions of well-being and the freedom to pursue well-being [Sen 1992]. In his text he describes capability as the freedom or opportunities to achieve well-being, whereas 'functioning' (akin to performance in this thesis) is the 'achievement' (also described as "beings and doings"). Thus, the ASKc may be a measure of potential, whereas the ASKp may be a measure of expression of that potential as achievement of activities within the context of the child's usual environments.

One of the primary objectives of the current research was based on this information, and was to explore the magnitude and nature of the difference between capability and performance. This was to be achieved by repeating the comparison of the ASKc and ASKp with a larger sample, and further exploring the meaning of any differences (e.g., in terms of environmental factors that relate to the expression and detection of disability) through qualitative methods.
Before embarking on further empirical comparisons, however, the literature relating to the impact of environment on disability was reviewed.

**Arguments for Environmental Role in Disability**

This review of the literature found many arguments supporting the concept that the environment, or the context in which physical disability is evaluated, is of vital importance in disability measurement [Barris 1987, Cott et al. 1995, Haley et al. 1994a, Kielhofner 1993, Law 1991, Law 1992, Law and Dunn 1993, Letts et al. 1994, Spencer 1993, Wilson and Cleary 1995]. There are three key conceptual arguments in favor of this position.

**First**, no human activity occurs in an absolute isolation (e.g., in an environment devoid of any other factors or influences). This has been stated explicitly by some: “no physical activity or task is strictly devoid of environmental factors or social context” [page 444 Haley et al. 1994a]. Other authors have confirmed the concept less overtly. Examples include, Cott et al.’s statement that “movement is also influenced by physical and social factors external to the individual” [page 89 Cott et al. 1995]. Kielhofner proposes a *dialectical evaluation* that produces “information about the person-environmental interface and that locates problems [disability] in the interface” [page 250 Kielhofner 1993]. Law’s qualitative research with parents of children with disabilities confirms this interactional model of person-environment [Law 1992]. Barris describes how “people, especially the elderly, will modify their activities in response to perceiving the environment in certain ways” [page 42 Barris 1987]. These observations are frequently and repeatedly confirmed in clinical practice. When individuals are asked about their abilities on a given activity (within the ICIDH disability domain) the answer is often “it depends upon the circumstances”.

The **second** argument in support of the impact of environmental factors on disability is that predominant models of disablement, specifically the ICIDH model [WHO 1980] and Nagi’s model [1965], both include an aspect of environment in their conceptualization [Haley et al. 1994a]. Although it is clear in the ICIDH model that the environment’s influence is expected to
be more important at the handicap level, it is not excluded from the disability level. For example, the ICIDH description of disability stipulates that it is the objectification of impairment, where objectification is defined as “the process through which a functional limitation expresses itself as a reality in everyday life” [WHO 1980]. Thus, the translation of impairment into disability is dependent upon the everyday environment in which the performance is expected. Recent discussions of the ICIDH model by both Verbrugge and Jette [1994] and Kopec [1995] have recommended making the role of environment at the disability level more explicit. Verbrugge and Jette suggest this by classifying some activities as situation-dependent [Verbrugge and Jette 1994], while Kopec presents an “Activity Space model” that suggests the actual performance of an activity is a function of impairment combined with the external circumstances and psychological factors [Kopec 1995].

The third argument in support of the importance of environmental factors on disability is the recognition of environmental factors in clinical practice models in rehabilitation medicine. This is evident in the Movement Continuum model [Cott et al. 1995] and the Occupational Performance model [Townsend et al. 1990]. The Movement Continuum model, created by a group of Canadian Physical Therapists in 1995, is a conceptual model of movement disorders that explicitly includes the impact of external environmental factors at all levels of movement disorders (e.g., impairment, disability and handicap) [Cott et al. 1995]. Some recognition of an environmental component to disability was also included in the model of Occupational Performance created by the Canadian Association of Occupational Therapist in 1983 (i.e., with regard to self-care activities) [Townsend et al. 1990]. Therefore, the concept that disability is altered by environmental factors is consistent with current theory in the Physical Therapy and Occupational Therapy professions.

Evidence Supporting the Adoption of an Environmental Role

The argument that activity does not occur in isolation, is supported by many studies demonstrating that disability scores change when the assessment environment is altered. For
example, patients' levels of disability observed in hospital do not reflect those observed at home immediately after discharge [Haworth and Hollins 1979, Park et al. 1993, Williams 1987]. The same changes have been reported for self-report instruments where the imagined reference environment is changed. For example, empirical comparisons of capability and performance have shown a statistically significant difference between these two measurement methods [Anderson et al. 1977, Young 1994]. Because capability questions imply a hypothetical set of circumstances, while performance questions specify the environment in which the subject has functioned recently, the difference between the two relates to a difference in context (hypothetical vs. real) [Spencer 1993, Young et al. 1996]. Thus, whether the environment is actually being experienced by the subject, or exists only in their mind, it appears to alter disability.

The result of this recognition of person-environment interaction is apparent in the discussion of disability models in the physical rehabilitation literature. There has been a call for replacement of models based on diagnosis and intervention, because these models do not accommodate chronic disability well [Law and Dunn 1993]. The proposals have been for new models referred to as ecological models [Albrecht 1992, Haley et al. 1994a, Law 1991] or socio-political models [Law and Dunn 1993]. Unlike the traditional medical model (that views the patient as a passive participant in the recovery process) [Albrecht 1992], these models propose options for treating not only the person with the disability, but also the environment that contributes to the disability. The Movement Continuum model [Cott et al. 1995] and the Occupational Performance model [Townsend et al. 1990] discussed previously, mark progress in this direction. However, what these models do not sufficiently emphasize, is that the person-environment interaction is not a static model, but is a dynamic model that is constantly in a state of flux [Letts et al. 1994]. Thus, it must be recognized that environmental changes may result in changes in the expression of disability.

The breadth of acceptance of the assumption that environment plays an important role in disablement has also been evident in its promotion by many disabled populations (e.g.,
independent living movement). These groups have used the relationship of person and environment for political purposes. They have used the premise that “the environment is at least as powerful a determinant of the functioning of people with disabilities as the capabilities of the individual” [Spencer 1993] as the basis for the independent living movement and the push for integration or “mainstreaming” of children with disabilities into the regular school and extra curricular activity programs.

Much of what is known about the process of environmental influences on disability has been generated from qualitative research (e.g., [Barris 1987, Law 1992, Williams 1987]. The concept of environmental effects on disability is commonly referred to as person-environment interaction [Letts et al. 1994]. Few specifics of the person-environment interaction have been identified to date. Law identified the following aspects of disabling environments: the built environment, societal allocation of space, classification based on norms, notion of disability as deviance, and the nature of the power structure within the health disciplines [Law 1991]. However, it appears that these factors may relate more to handicap than to disability. Thus, relatively little is known about specific aspects of the environment that contribute to physical disability.

The search for factors specific to physical disability in children may be facilitated by information from other fields. Letts et al. point out that person-environment models are common to other professions such as planning, architecture, psychology, gerontology and human geography [Letts et al. 1994]. Theorists in these fields have postulated that it is not just the environment, but rather the person’s perception of that environment that is important [Baker and Intagliata 1982]. Others have postulated that it is the social environment [Kahana 1982] or physical environment [Weisman 1981] that is more important. Barris specifies that although perception alters environment, participation may alter perception. Thus by changing one’s perception of an environment to a more favorable one based on experience, disability may be reduced [Barris 1987].
The preliminary qualitative information obtained in my MSc research [Young 1994] empirically support the second argument, that environment belongs in conceptual models of disability. The interviews, conducted for the purpose of item generation for the ASK, identified factors beyond the person level that impacted on children’s disability. For example, disability in walking was reported to depend both on the abilities of the child and the characteristics of the ground. Thus, disability may be situation-dependent, like handicap, although perhaps to a lesser degree.

The preceding paragraphs have outlined the conceptual basis, empirical evidence, and broad acceptance of the role of the environment in disability. However, there has been little information on environmental factors included in disability measures [Young and Wright 1995]. Spencer suggests that this may be due to the fact that the evaluators of disability are clinicians whose contact with patients has been almost exclusively within the institution, and because of clinician’s preference for quantitative measurement [Spencer 1993]. Quantitative methods are limited in their ability to summarize environmental information. Thus, qualitative methods may be more appropriate for the identification of environmental factors and the exploration of how they are integrated into disability.

In summary, the conceptualization of disability in this thesis is one in which disability is a composite of a person’s ability and environmental factors. This conceptualization will be tested and enhanced as part of the thesis.

**PROPOSED WORK**

The literature on measurement issues, existing scales, and the ASK forms the basis for new research to evaluate the measurement properties of the ASK. The literature on disability models and environmental impact forms the basis for the new research to explore the difference between capability and performance and thereby improve the conceptualization of childhood disability.
CHAPTER 3

METHODS

The review of literature highlighted the need for a physical disability measure for five to 15 year old children with musculoskeletal disorders, and the importance of such a measure for evaluating the effectiveness of treatments. The clinimetric development of the Activities Scale for Kids (ASK) was documented as well. The ASK is considered to have good content validity by virtue of the clinimetric methods used. The previous chapter also recounted the “excellent” [Landis and Koch 1977] reliability of both versions of the ASK, and introduced the importance of considering environmental context in measuring children’s disabilities.

OVERVIEW OF METHODS

The primary objectives of this study were: further evaluation of the measurement properties of the ASK, and the exploration of the contextual factors responsible for the differences between capability and performance limitations in five to 15 year old children. These two primary objectives were addressed by a series of eight sub-studies. An overview of these methods is shown in Table 4 on the next page. Essentially the present research was composed of one main baseline data collection from 200 children. This sample size was based on the component with the largest sample size requirement (item characteristic analysis), and was more than adequate for the additional comparisons. From this total sample, mutually exclusive sub-groups were identified for additional testing: 24 were required to attend a clinical assessment, 30 or more were required to completed questionnaires after clinically important change, and up to 40 were expected to participate in individual interviews. The purposes of each of these sub-samples are summarized in Table 4 and in Figure 1 (page 81).
# Table 4 Summary of Methods

<table>
<thead>
<tr>
<th>Primary Objective</th>
<th>Sample Design</th>
<th>Total Sample n=200</th>
<th>Sub-Sample I n=24</th>
<th>Sub-Sample II n=30</th>
<th>Sub-Sample III n≤40</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample</strong></td>
<td></td>
<td>Baseline Assessment</td>
<td>Clinical Assessment</td>
<td>Change Assessment</td>
<td>Interview</td>
</tr>
<tr>
<td><strong>Data Collected</strong></td>
<td></td>
<td>ASKp and ASKc from children</td>
<td>ASKc-observational collected from each child and 2 clinicians independently</td>
<td>ASKp and ASKc from children</td>
<td>Interview tapes from children</td>
</tr>
<tr>
<td><strong>Assessment of ASK's Measurement Properties</strong></td>
<td><strong>Item Characteristic Analysis</strong></td>
<td>Rasch Modeling confidence limits of ±0.1 logit (SS1)*</td>
<td>Criterion Validity II (discriminative validity) relative to clinicians' global ratings (SS3)</td>
<td>Criterion Validity I relative to clinician-reported ASK based on observation (SS2)</td>
<td>Construct Validity III (Responsiveness) comparison to clinical change and calculation of effect sizes (SS6)</td>
</tr>
<tr>
<td><strong>Validity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Construct Validity I &amp; II</strong></td>
<td>relative to CHAQ and HUI3 (SS4 &amp; SS5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Exploring the Determinants of Physical Disability</strong></td>
<td><strong>Comparison of Capability and Performance</strong></td>
<td>Statistical Comparison ASKp &amp; ASKc based on binary and percentile scores (SS7)</td>
<td></td>
<td></td>
<td>Qualitative Analysis of factors that mediate between capability and performance (SS8)</td>
</tr>
</tbody>
</table>

* Sub-Study number, corresponds to the sub-study numbers shown in Figure 1.
A full explanation of the rationale and sample size calculations for each component of the study’s methods are described in this chapter.

**Eligibility**

Eligible subjects (hereafter referred to as children) included all children 5.0 to 15.99 years of age, who were experiencing either acute or chronic limitations in physical activity due to musculoskeletal disorders (i.e., arthritis, fracture, spina bifida, etc.). Those who had impaired cognitive status (defined as reading below grade three level if older than eight years of age, or below age-expected cognition if younger than eight years of age) were excluded due to the requirement of self-report. These children could be appropriately evaluated by parent-report, using either the ASK (previously shown to be concordant with child-reported ASK scores [Young 1994]) or the Pediatric Evaluation of Disability Inventory (PEDi), or by clinicians using the Gross Motor Function Measure (GMFM).

**Recruitment Process**

Although the present research had eight sub-studies, there was significant overlap in the data required for the analyses of each sub-study. Therefore, there was one main recruitment drive to enroll patients into the entire study and to collect the baseline data necessary for many of the sub-studies (see second row of Table 4 for clarification of the data required for each sub-study).

Children with physical disabilities were identified by clinicians in the Orthopaedic, Physical Therapy, and Rheumatology clinics at The Hospital for Sick Children (HSC), The Bloorview MacMillan Centre (BMC), Erinoak (EO), Variety Village (VV), and from the Children’s Hospital of Michigan (CHM). HSC was selected to provide referrals from an acute care facility. BMC was selected to provide referrals from a rehabilitation facility. HSC and BMC were expected to provide the majority of referrals. However, EO and VV were included to access children in the community who may not be actively involved with either HSC or BMC. The last centre, CHM, was added later in order to recruit more children with
orthopaedic conditions who were expected to change in a predictable direction, and could contribute to the responsiveness analysis. Ethics approvals were obtained from HSC, BMC and CHM, and are included in Appendix V. EO and VV did not have ethics committees, and deferred review to HSC.

Several strategies were used to ensure a broad range of children were recruited. In some cases this was facilitated by accessing clinical data bases (e.g., Spina Bifida program at BMC), while in others a review of clinic charts and consultation with medical staff was used (e.g., Orthopaedic clinics at HSC). The sampling strategy focused on recruiting children with a diverse range of ages, disability levels (based on referring clinician’s global assessment) and diagnoses to enhance the generalizability of the results. This was achieved by monitoring the distribution of children and encouraging the preferential recruitment of children from specific age, disability level or diagnostic groups when necessary.

Consent Methods

Parents of eligible children identified through these clinics were given a letter describing the study along with copies of the information, parental consent and child assent forms (shown in Appendix V). Families were later asked verbally if they wished to participate. Those who verbally consented were enrolled, and written consent and assent were obtained in the first round of data collection. Although the consent process covered both the initial data collection as well as all potential additional components of the study, verbal confirmation of continued consent was obtained for all children before entering them into a second component of the study (e.g., responsiveness assessment).

PART A: ASK MEASUREMENT PROPERTIES

The first of the two primary objectives of the study was to evaluate the ASK to ensure that it had appropriate item characteristics, and was valid, responsive, and clinically useful. This
required evaluation of item characteristics, and scale properties of validity and responsiveness. All 200 children participated in the initial data collection session. In addition, some children subsequently participated in a second component. However, no child completed more than one additional component.

Reliability had previously been assessed and shown to be "excellent" [Landis and Koch 1977] (intraclass correlation 0.97) [Young et al. 1995]. However, the previous assessment of validity had been coarse, in that it assessed the ASK's ability to differentiate between three levels of disability (mild, moderate and severe) [Young 1994], and responsiveness had not previously been assessed. Therefore both of these properties were addressed in the present research. The methods specific to the evaluation of each of these properties are presented in the following sections. The methods have been organized into a series of sub-studies. Thus, the first sub-study's rationale, sample size calculation, data collection and analysis components are presented together, followed by the same information for the second sub-study, etc. A synopsis of the series of sub-studies is presented in Table 4 (see page 49).

ITEM CHARACTERISTIC ANALYSIS

The purpose of this component of the research was to identify the structure of the item content so that future refinements of the ASK could enhance feasibility (by selecting a smaller item set) while continuing to optimize the measurement properties of the ASK. In its 53 item format, each version of the ASK required approximately 30 to 45 minutes to complete the first time. The attention span of children and the preferences of clinicians both identified approximately 20 minutes as the ideal length of time to complete a disability questionnaire. Thus, the ASK's items needed to be reviewed and the number reduced, provided this was possible without compromising its measurement properties.

The identification of the ASK's item structure was an essential aspect of this process. The term item structure [Nunnally and Bernstein 1994] is used to encompass a description of the dimension measured, the range of the item set, the spacing of the items, the specific location of
individual items across the dimension, and the degree of confidence around that location. For example the structure of a house can be described in terms of the vertical dimension, having 160 bricks covering a five metre range, equally spaced, with one brick of interest located three metres from the base ± 0.01 metres. The intent of item structure assessment was to provide a similar description for the ASK’s items, so that both the length of the ASK and its measurement properties could be optimized by selecting those items that were unidimensional and equidistantly spaced across the full spectrum of physical disability [Nunnally and Bernstein 1994]. This information on item structure was later used to make recommendations regarding revisions to the item content of the ASK, in concert with information on items’ reliability, validity and responsiveness (see Appendix III).

Item characteristic analysis (ICA) is the most appropriate means of identifying a scale’s item structure [Nunnally 1978]. Other methods (e.g., factor analysis) may identify the affinity of certain groups of items to specific dimensions, but cannot pinpoint the location of one item relative to another within one dimension [Nunnally and Bernstein 1994]. Knowledge of item location is critical to the selection of equidistant items and the elimination of redundancy, thus was the primary reason for selecting an ICA method as the keystone of the assessment of item behaviour.

Selection of an Item Characteristic Analysis Model

ICA models are distinguished from one another by the number of parameters they contain, ranging from one to three (described in detail in Appendix II). The model selected for the present research was a one-parameter model known as the Rasch model [Hambleton 1991]. It was selected because it provided the item locations and an excellent selection of diagnostic indicators related to item behaviour (i.e., fit statistics) [Linacre and Wright 1993]. The Rasch model also satisfies the aim of freedom from sample bias and freedom from item selection bias

---

1 Freedom from item selection bias means that the scores of individuals are not altered by the specific subset of items administered (or answered) from the larger pool of items on the questionnaire.
that cannot be met adequately by the alternative approaches [Fisher et al. 1995, Molenaar 1995] (discussed in the review of literature under “Methods of Scale Development” and in Appendix II). For these reasons Nunnally, in 1978, clearly stated that Rasch modeling would be the preferred method if it were more easily performed [Nunnally 1978]. At that time, programs such as the Bigsteps® program [Linacre and Wright 1993, Wright and Linacre 1992] selected for this work were in their infancy. However, programs have subsequently become much more practical and can be run on personal computers at minimal cost [Nunnally and Bernstein 1994]. The Rasch model selected for the present research has previously been applied to several other measures of health, including the Functional Independence Measure [Fisher et al. 1995, Granger et al. 1989], SF36 [McHorney 1994], and PEDI [Haley et al. 1992]. Molenaar summarizes the advantages of item characteristic analysis or item response theory (IRT) as outweighing previous approaches: “IRT can do the same things better and can do more things, when it comes to modeling existing tests, constructing new ones, applying tests in non-standard settings, and above all interpreting the results of measurements” [page 4 Molenaar 1995].

Description of Rasch Analysis

The one-parameter Rasch model has three primary assumptions: (1) unidimensionality, (2) local independence of items, and (3) sample independence [Hambleton 1991]. Unidimensionality refers to the items all belonging to a single dominant domain (i.e., physical disability) [Haley et al. 1994b, Tennant 1995]. When this assumption is not true in a given data set, items from the non-dominant domain have high infit and outfit statistics (described in Appendix II) because the items from an opposing domain do not vary according to the pattern predicted for the domain [Andiel 1995, Wright and Linacre 1989]. Local independence refers to all items measuring a unique aspect of the domain, distinct in some subtle way from all other items [Hambleton 1991]. When this does not hold true, items have low infit and outfit statistics because there is redundancy relative to other items [Linacre and Wright 1993, Linacre
Sample independence refers to the assumption that the results of modeling will not be influenced by the particular sample of the population chosen [Fisher et al. 1995, Hambleton 1991]. If this assumption were not true, then washing would be modeled as more difficult than washing oneself in one sample of children and as less difficult in another sample. This is evident if there are sub-groups of children with very high infit or outfit statistics [Linacre 1996.

The Rasch model evaluates item structure by an unconditional maximum likelihood estimation process (referred to as the ‘American Tradition’ as opposed to the conditional ‘European Tradition’) that attempts to optimize differences between children’s summary scores and items’ mean scores simultaneously [Linacre and Wright 1993, Tennant 1995]. This is accomplished by comparing response patterns observed across children for individual items to response patterns observed across items for individual children, using an iterative process. The modeling process is based on three key premises: (1) that persons of high ability should achieve most of the items (except the most difficult items) while persons of low ability should achieve only the easiest items, (2) that hard items are achieved by only the most able persons while easy items are completed by most persons, and (3) that variability exists in response patterns at the point of transition from able to unable [Hambleton 1991].

As a result, Rasch analysis describes the behaviour of items and persons so as to maximize the consistency of response patterns [Hambleton 1991]. The behaviour of each response option for each item is plotted as the probability of success (y-axis) against the difference between the child’s ability and the items demand (x-axis), and referred to as an item characteristic curve or item response function (IRF). When an item has only two response options, Rasch analysis models the probability of a positive response (odds of scoring one as opposed to zero) and the shape of the IRF is a logistic function (similar to the cumulative normal distribution). When an item has three or more responses, the probability of each response may be plotted. In this event the probability of the third response option is a logistic function, the probability of the second response approximates a normal distribution, and the
probability of the first response option is a mirror image of the logistic function of the third response option. Alternatively only the median response option IRF is presented, and the point along the x-axis at which that median IRF crosses the 50% probability point on the (y-axis) is the difficulty level of the item expressed in log-odds units (logits). Positive logit values denote items that are easy to complete for the average child in the reference sample and negative logits denote items that are hard for the average child, because the mid point of the x-axis is usually assigned the value 0. These logit values describe the location of the ASK items. A similar graph of a child’s responses across all items is used to describe the child’s ability level in logits. The range of logit values that describe the locations of items and children along their respective disability continuums is usually between negative two and positive two but may vary as widely as negative four to positive four [Linacre 1996].

Rasch analysis offers several additional advantages. It is able to transform the raw scores to linear measures so that parametric analysis methods may be used [Andiel 1995, Wright and Linacre 1989]. It is also able to test the stability of item parameters between different versions (e.g. ASK-capability and ASK-performance versions) or diagnostic groups, and therefore assesses the degree of “differential item functioning” (the degree to which the item behaviour is altered by differences in the sample population) [Groenvold et al. 1995] or item bias [Linacre 1996, Nunnally and Bernstein 1994]. It handles missing data without loss of power [Fisher et al. 1995, Linacre 1996]. Finally, it is useful in preparing for the possibility of diagnosis specific tailored tests and item banking in the future [Linacre 1996, Nunnally and Bernstein 1994].

In contrast to the advantages, the disadvantages must also be considered. There are four main disadvantages [Hambleton 1991, Linacre 1996]. First, the model is unstable with sample sizes less than 25 [Linacre 1994]. This is not substantially different from linear modeling or factor analysis, but may be problematic with rare clinical populations. Second, the Rasch model is a constrained model, therefore the fit of the data to the model is assessed rather than the fit of the model to the data. Unlike factor analysis, the model cannot be manipulated to
improve fit, however, the data may be reviewed and cleaned by the elimination of misfitting items to improve fit to the model [Molenaar 1995]. Third, the nature of the model prohibits the inclusion of hierarchically chained items [Streiner and Norman 1995] (e.g., climbing one flight of stairs, climbing two flights of stairs), and the model will report misfit if such items are included. This is likely a necessary restriction for most disability measures. Fourth, the requirement of invariance of models across diverse samples may be difficult to meet in practical application, because no item will fit the model exactly [Streiner and Norman 1995]. Thus, the degree of variance may be more relevant than its presence or absence. However, despite these disadvantages, the Rasch model was considered more appropriate than other methods for determining the item characteristics of the ASK and diagnosing regions of misfit within the ASK.

**Sample Size**

The sample size for the Rasch analysis was based on the number required to provide narrow confidence intervals around the item location parameters. Linacre suggests that deviations of up to 0.5 logits are acceptable, and that for item parameter estimates to be stable within 0.5 logits at 99% confidence, 150 children are required [Linacre 1994]. However, a sample of 50 subjects is usually sufficient if subjects answer more than 10 items [Wright 1996]. To meet the sampling needs of Rasch analysis and all sub-components, a sample size of at least 150 was required. This sample was revised to 200 to meet the more conservative sample size estimations presented in general measurement texts [Nunnally and Bernstein 1994, Streiner and Norman 1989].

Although the framework of the Rasch model is theoretically not sensitive to differences in population, inclusion of both clinical and community samples is an asset to generalizability [Molenaar 1995]. This range was facilitated by including a mix of clinical and community recruitment centres.
Data Collection

Referring clinicians were requested to provide a global rating of the child's disability level. Because clinicians have differing definitions of disability it was necessary to standardize their global assessments. Clinicians were, therefore, requested to answer five standard questions about the child's abilities. Based on their answers and an algorithm devised by clinical consensus, each child was assigned a global rating of mild, moderate or severe (see Appendix VI).

During the initial data collection session, ASKp (green) and ASKc (yellow) questionnaires (shown in Appendix VII) were administered by mail to all 200 children with physical disabilities. These children belonged to families who had verbally consented to participate. This data collection package also contained information forms, consent forms, CHAQ and HUI3 forms (the HUI3 has been reproduced with permission in Appendix VIII) for parents to complete, and a postage paid envelope to return the forms. All mailed materials for children were printed on coloured paper, and all materials for parents were printed on white paper for ease of administration. Note that the CHAQ and a HUI3 forms were not necessary for this particular section of the methods but were included in this package for use in other sub-studies.

The ASK versions were answered by the children with the provision that parents could read the questions to children under the age of nine (however, the children still determined and recorded the answers). The completed forms were returned in postage-paid envelopes. Reminder cards were sent to those who did not respond after one month, and follow-up phone calls were made after two months.

Analysis

All data were double entered and cleaned prior to the analysis to eliminate data entry errors from affecting the analyses. Summary scores were calculated for the ASK versions by averaging the answers to all completed questions and the results were reported as a percentage score, where 100% indicated full physical function. Note that this aggregation method
effectively eliminates missing values and not-applicable responses (blank cells) from both the numerator and denominator of the summary score calculation, and is consistent with the manner in which missing values and not-applicable responses are dealt with in the CHAQ. Thus, the ASK scores represent how well the children did on all applicable items. A list of the domains with sample items and scoring codes is shown in Table 5. Complete copies of the ASK questionnaires and details of the formulae for the summary score calculations are presented in Appendices VII and IX respectively. In cases where two sets of data were available for one child (responsiveness sample of n=30), the data from the first time period (baseline) was entered into the main data base for consistency.

Table 5 ASK Domains and Sample Items

<table>
<thead>
<tr>
<th>Number of Items Per Domain</th>
<th>Domain</th>
<th>Sample Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Personal Care</td>
<td>turned on the tap</td>
</tr>
<tr>
<td>2</td>
<td>Washing</td>
<td>washed my whole body</td>
</tr>
<tr>
<td>2</td>
<td>Toileting</td>
<td>got on and off the toilet</td>
</tr>
<tr>
<td>6</td>
<td>Dressing</td>
<td>put my shirt on</td>
</tr>
<tr>
<td>3</td>
<td>Transportation</td>
<td>got on and off a bus</td>
</tr>
<tr>
<td>2</td>
<td>Eating and Drinking</td>
<td>ate my meals</td>
</tr>
<tr>
<td>2</td>
<td>Food Preparation</td>
<td>made a snack</td>
</tr>
<tr>
<td>1</td>
<td>Miscellaneous</td>
<td>took care of my medical needs</td>
</tr>
<tr>
<td>9</td>
<td>Walking and Rolling</td>
<td>got around inside my home</td>
</tr>
<tr>
<td>3</td>
<td>Carrying</td>
<td>carried my backpack</td>
</tr>
<tr>
<td>3</td>
<td>Stairs</td>
<td>climbed up and down a step or curb</td>
</tr>
<tr>
<td>6</td>
<td>Spare Time Activities</td>
<td>ran around and played outside</td>
</tr>
<tr>
<td>3</td>
<td>Sitting</td>
<td>sat on the floor</td>
</tr>
<tr>
<td>3</td>
<td>Transitions</td>
<td>got in and out of my bed</td>
</tr>
<tr>
<td>4</td>
<td>Standing</td>
<td>stood for 10 minutes</td>
</tr>
<tr>
<td>1</td>
<td>Entering/Exiting Buildings</td>
<td>got through heavy doors</td>
</tr>
</tbody>
</table>

*ASKp Response Options*  
4 = all of the time  
3 = most of the time  
2 = sometimes  
1 = once in a while  
0 = none of the time  
n/a = not-applicable  

*ASKc Response Options*  
4 = with no problem  
3 = with a little problem  
2 = with a moderate problem  
1 = with a big problem  
0 = I could not
Descriptive statistics were then calculated to determine whether the distributions of disability scores from the various measures were close enough to normal (as defined by normality tests [Hintze 1992]) to use parametric as opposed to non-parametric statistics, and to describe the population. Descriptive statistics were also used to assess the frequency of missing values and not-applicable responses, and inferential statistics were used to assess the potential impact of these blank cells. Note that because the ASKp was intended to measure disability in one's usual environment it contained not-applicable response options, whereas the ASKc did not. If missing values and not-applicable responses accounted for more than 20% of the data on any item, then alternative aggregation methods would be explored to determine whether or not systematic bias had been introduced by the blank cells.

The item structure of the ASK was identified using Rasch modeling via the Bigsteps® computer program, beginning with a partial credit model. In the partial credit model each item is assumed to have a unique response set (i.e., response option A may be twice as easy as response option B for item two and A may be four times as easy as B for item two). If the ordering of responses and distance between response options is consistent across items, then the partial credit model need not be used in subsequent analyses. Its early use, however, is important in documenting the appropriateness of the single rating scale model, that assumes one standard response set across items. This analysis was run on the combined data set from all 200 children gathered during the initial data collection session.

**VALIDATION STUDIES**

The validity of the ASK was systematically tested against several standards according to the ascending purposes of measurement described by Kane and Kane [1981] and summarized in Table 1. The basic prerequisites for a descriptive measure are that it have face validity, internal consistency and preferably content validity [Kane and Kane 1981]. These parameters were achieved by virtue of the ASK’s clinimetric development methods, that have been previously documented [Young 1994] and were summarized in Chapter 2, Table 3.
To be useful at the screening level (i.e., applied to explore large groups for potential disorders) the ASK must be able to discriminate between levels of disability, at least as well as clinicians, but by a more efficient means (i.e., self-administered questionnaire as opposed to observational assessment). Preliminary evidence suggested the ASK had this ability, but the standardized difference scores between children with mild and moderate disabilities were very small [Young 1994]. Thus, the assessment of discriminative validity was refined and repeated in this thesis.

**Criterion Validity I (comparison to clinician observation)**

At the monitoring level, proof of criterion validity and responsiveness with fine gradations becomes important. As previously mentioned, there were no existing questionnaires that provided an accepted external standard for monitoring change in physical disability in this population. The most commonly accepted standard for physical disability in children is clinician observation. Thus, criterion validity (defined as comparison to a concurrent or future external standard) was assessed by comparing child-reported ASK summary scores with clinician-reported ASK summary scores. The clinician's based their report on clinical observation.

For this comparison only the ASKc was employed, because this version was considered to be more consistent with the clinical environment in which the children would be assessed for criterion validity. The ASKp records the frequency of performance in the child's usual environment and the introduction of a clinician into this usual environment would render the environment unusual or artificial. Furthermore, it was not feasible for a clinician to observe for long enough periods of time (e.g., several days) to accurately assess frequency of performance, whereas capability could be more quickly assessed.

Three modifications were made to the ASKc for the purpose of criterion validity assessment. First, the five point ordinal response set was changed to a four point ordinal response set, to reflect the number of gradations in the capability that could accurately be
assessed by an outside observer (no problem, a little problem, a moderate problem, a big problem, unable). Second, only 30 of the items (57%) were considered to be observable in a clinical setting (e.g., toileting was inappropriate, and playing with friends was impossible). Thus, a modified version of the ASK (called the ASK-observational) was created that contained 30 items with a four point ordinal response set. Third, a random sample of 15 of the questions from the ASK-observational version were administered to each child once to keep the assessment under one hour and within the attention span of most children.

Sample Size for Sub-Sample I
Sample size for the clinical assessment sub-group was based on the hypothesis that the correlation between clinician and child-report would be approximately 0.8 and significantly greater than 0.5, with α=0.05. Cohen's formula for comparison of correlations [page 110 Cohen 1988] was used to estimate the sample size.

\[ q = z_1 - z_2 \text{ where } z = \frac{1}{2} \ln \frac{1 + r}{1 - r} \text{ then } q = 1.099 - 0.549 = 0.550 \]

Since this comparison was of a correlation from a sample compared to a criterion, value q is modified as follows: \( q' = q \sqrt{r} = 0.776 \). Based on Cohen's tables, this value of q' requires a sample size of 24 [page 134 table 4.4.1 Cohen 1988] to be certain of detecting a correlation of 0.5 or greater. Although this formula assumes normality of the sample distributions, "considerable departures" [page 109 Cohen 1988] from this assumption have little effect on the estimates.

This sample of 24 children with musculoskeletal disability was drawn from the larger study. The data for this section could not be obtained as part of the first session of data collection. The criterion validity component of the study required patients to travel to either HSC or BMC for clinical assessment. Therefore, it represented a large incremental demand on parents' time, and only a sub-set of the sample were requested to participate. These children represent a sub-sample of convenience (taken from an initial sample of 200), in that they were restricted to children who either lived within one hour of Toronto, or who would be in the city
for treatment. An attempt was made to ensure representation across all age groups, and an equal gender distribution.

**Data Collection**

Children who were enrolled in the study, and were either completing or had completed their first data collection package by mail, were requested to attend a clinical assessment. During this assessment the child first answered a random sample of 15 ASK-observational version items. The child then demonstrated each of the 15 items in front of two clinicians (A and B) blind to the child’s responses. Each clinician independently rated the child’s physical function using the ASK-observational version. After completing the activities the child was given a blue pen and permitted to remark any item he or she felt they had demonstrated at a level higher or lower than they had reported in advance of attempting the activity. The purpose of this final review of answers was to determine whether children’s perceptions changed after physically trying the activity. This information would be analyzed to determine the source of disagreement in the event that child-report was discordant from clinician-report.

**Analysis**

Summary scores were analyzed as previously described, however, the maximum number of items was 30. The primary analysis was conducted using the children’s first ratings compared to the clinician A’s ratings. Correlational methods were used to determine the strength of the relationship between child-report and clinician-report. Correlation was assessed rather than concordance because some difference between clinicians and children’s perspectives was expected. In order to generalize the results from ASK-observational scores to ASKp and ASKc scores, correlations between the ASK-observational, the ASKc and the ASKp were run. Intraclass correlation coefficients were calculated to determine the inter-rater reliability between clinicians A and B. Exploratory analyses were performed to assess the concordance between child and clinician, the effect of post-demonstration ratings by the children, and the validity
Criterion Validity II (comparison to global ratings of disability)

Evaluation of the discriminant form of criterion validity required comparison of ASK scores between three groups known to be clinically different. Previous research had suggested it was most difficult to discriminate between children with mild and moderate disability. This difficulty was attributed to two sources. First, the standard for assigning children to mild, moderate and severely disabled groups was based on a clinician’s general impression of the child’s ability, rather than a more structured report. Secondly, at the time of that research the ASK contained 73 items, many of which were not detecting disability in any of the children, thus were diluting the difference in scores between groups. These two deficiencies were addressed by creating a more structured approach to global ratings based on clinician consensus, and by the preliminary item reduction strategy described in Appendix III (Item History).

Sample Size

The sample size estimation was based on an independent groups t test for detecting a statistically significant difference between two groups of children: those with mild disability and those with moderate disability. This calculation [page 84 Meinert and Tonascia 1986] employs the concept of effect size that was defined as the difference in scores between groups divided by the pooled standard deviation [Cohen 1988, Lipsey 1990]. Data from the ASK’s initial evaluation on 28 children showed the effect size between mild and moderately disabled groups was 0.32 for the ASKp summary scores (difference=d=1.55, variance=var=4.85) and 0.38 for the ASKc summary scores (d=3.05, var=7.98). However, improvements have been made in both the scale and the classification of children since that study. Therefore, the sample size calculation was modified to ensure the ability to detect a difference of 0.5 standard deviation units when α=0.05 and β=0.2.
Thus, 63 children with mild disabilities and 63 children with moderate disabilities were required to detect the difference between mild and moderately disability groups [Mei\-nert and Tonascia 1986].

The effect size previously observed between moderate and severe groups was 3.03 for the ASKp summary scores (d=.28.62, var=9.46) and 3.54 for the ASKc summary scores (d=.29.64, var=.836). However, the improvements made to the global rating process to enhance discrimination between mild and moderate groups may have attenuated the difference between moderate and severe groups. Thus, the sample size calculation was calculated conservatively, to ensure the ability to detect a difference of 2.0 standard deviation units when \( \alpha=0.05 \) and \( \beta=0.2 \). Because the predicted difference is extremely large, application of the previous formula concluded that four children with severe disabilities were required to detect the difference between moderate and severe disability groups [Meinert and Tonascia 1986].

\[
\text{Effect Size} = \frac{15.68}{0.5^2} = 62.7
\]

\[
n_{\text{min}} = n_{\text{mod}} = \frac{2(Z_\alpha + Z_\beta)^2 \text{var}}{(\mu_\text{mod} - \mu_\text{sev})^2} = \frac{2(1.96 + 0.84)^2 \text{var}}{d^2} = 15.68 \times \text{var} = 15.68 \times d^2 = 62.7
\]

Thus, 63 children with mild disabilities and 63 children with moderate disabilities were required to detect the difference between mild and moderately disability groups [Meinert and Tonascia 1986].

The effect size previously observed between moderate and severe groups was 3.03 for the ASKp summary scores (d=.28.62, var=.9.46) and 3.54 for the ASKc summary scores (d=.29.64, var=.836). However, the improvements made to the global rating process to enhance discrimination between mild and moderate groups may have attenuated the difference between moderate and severe groups. Thus, the sample size calculation was calculated conservatively, to ensure the ability to detect a difference of 2.0 standard deviation units when \( \alpha=0.05 \) and \( \beta=0.2 \). Because the predicted difference is extremely large, application of the previous formula concluded that four children with severe disabilities were required to detect the difference between moderate and severe disability groups [Meinert and Tonascia 1986].

\[
\text{Effect Size} = \frac{15.68}{2.0^2} = 3.92
\]

Given that the sum of these estimated sample sizes were less than the 200 children who were required for item characteristic analysis, an attempt was made to ensure that the total sample met the severity distributions required for the discriminative validity assessment. Thus, there was intentional over sampling of children with mild and moderate disabilities, as shown in Figure 1 (\(~115\) mild + \(~115\) moderate + \(~56\) severe). This strategy was intended to more accurately approximate the distribution in the population, while ensuring sufficient numbers in the smallest group. There are very few children with severe disabilities (4%), relative to the numbers of children with mild (85%) and moderate (11%) disabilities in the community [Hanvey et al. 1994a].
Data Collection

Data were collected on the ASKp and ASKc, as previously described, from 200 children. No additional data were required for the purpose of comparing ASKp scores to global ratings of disability severity.

Analysis

Summary scores were calculated as previously described (see also Appendix IX). The ability of the ASKp and ASKc to detect differences between children with mild, moderate and severe disabilities, as indicated by these global ratings, was determined by one-way analysis of variance. If a statistically significant difference was found between groups, this analysis was followed by independent groups t test comparing mild to moderate and moderate to severe groups.

Construct Validity I (comparison to CHAQ)

In addition to the assessment of criterion validity, several forms of construct validity were also assessed. First, the construct validity of the ASK was directly evaluated by comparing the ASK to a similar physical disability measure. The most relevant disability measures were reviewed in Chapter 2. Among these measures, the Childhood Health Assessment Questionnaire (CHAQ) was considered to be the most appropriate for this study’s musculoskeletal population.

Although the CHAQ is disease specific, it has been used clinically with other diagnostic groups (e.g., dermatomyositis[Feldman et al. 1995]). Thus, it was thought to provide a reasonable estimate of disability levels. However, because of the limitations in the CHAQ discussed in Chapter 2 (it is specific to juvenile rheumatoid arthritis, is a modification of an adult measure, lacks sensibility in the aggregation methods, and has potential for confounding of age and disability effects) it was not considered to represent an accepted standard for children with a broad range of musculoskeletal conditions. Thus, comparisons between the
ASK and the CHAQ were not classified as an assessment of criterion validity. Rather, the comparisons between the ASK and CHAQ were classified as an assessment of construct validity, where the common construct was physical disability as defined by the international classification of the consequences of disease [WHO 1980].

**Power Calculation**

The sample size calculation for construct validity of the ASK relative to the CHAQ was based on proving that the correlation between the two scales was significantly greater than 0.5 with the assumption that the true correlation was approximately 0.8. The formula previously described for criterion validity I (versus clinical observation) was used to estimate the sample size. The resulting sample size estimate was 24 [Cohen 1988]. However, all 200 children in the first data collection session were included to improve the accuracy of the point estimate, resulting in the power of the test exceeding 0.99.

**Data Collection**

In addition to the ASK, global rating, and HUI3 data collection already described, parent-reported CHAQ data were also collected in the initial data collection session, so that construct validity could be further assessed. Parent-report was used on the CHAQ rather than child-report for two reasons: (1) the literature on the CHAQ has most often been based on parent-report data, and (2) a strong correlation with a second respondent would enhance the generalizability of the results.

**Analysis**

Aggregation of the ASK scores was previously described and is reviewed in Appendix IX. CHAQ scores were calculated in accordance with the developers instructions (sum of the maximum score in each of eight sub-domains with a minimum score of two for categories in which assistive devices were used or help was accepted divided by eight times three as shown in Appendix IX). The scores were then reversed and presented as a percentage score (to be
consistent with ASK scores), where 100% indicates full physical function. ASK scores were compared to CHAQ scores using Pearson’s correlations to evaluate construct validity.

**Construct Validity II (comparison to HUI3)**

A final form of construct validity was assessed relative to a screening measure: the Health Utilities Index Mark III (HUI3). The HUI3, described in Chapter 2, was an eight domain descriptive measure that has been used to collect demographic information on childhood disability in Canada. Two of the domains (ambulation and dexterity) were measures of physical disability, and two measured distinctly different constructs (emotion and speech). Although the level of discrimination within the ambulation and dexterity domains was considered to be insufficient to evaluate therapeutic effectiveness, it was considered to be sufficient for the assessment of convergent validity. Similarly, the level of discrimination within the emotion and speech domains was expected to be sufficient to assess the divergent validity of the ASK.

**Power Calculation**

To assess convergent and divergent validity, it was hypothesized that the correlation between the ASK and similar constructs would exceed 0.7 and be statistically greater than the correlation between the ASK and divergent constructs (that would not exceed 0.3). Cohen’s formula for comparison of correlations[ page 110 Cohen 1988] was used to estimate the sample size.

\[ q = z_1 - z_2 \quad \text{where} \quad z = \frac{1}{2} \ln \frac{1 + r}{1 - r} \quad \text{then} \quad q = 0.867 - 0.310 = 0.557 \]

Taken to tables, this value of q requires a sample size of 43. However, there data were collected as part of the initial data collection sample that had a base sample of 200 children. With a sample size of 200 the power of the test would exceed 0.99.


**Data Collection**

Therefore, in addition to the ASK\(_{p}\), ASK\(_{c}\), and global rating data collection already described (see previous section), HUI3 data were also collected from the parents of the 200 children with physical disabilities. Because HUI3 data are usually collected by interview rather than by self report, the HUI3 was collected by both self report and later by interview. This was performed with a sub-set of parents of physically stable children who were seen in person for the purpose of clinical assessment or interview.

**Analysis**

The analysis for this section employed correlations between two convergent and two divergent constructs. Each was correlated with the ASK\(_{p}\) using Spearman’s correlation. Spearman’s correlation was selected over Pearson’s because the HUI3 sub-domain scores are ordinal and have only six gradations, thus do not approximate a continuous measure. A “high” correlation [page 94 Nunnally and Bernstein 1994] between the ASK\(_{p}\) and HUI3 ambulation and dexterity sub-scales coupled with a “low” correlation [page 94 Nunnally and Bernstein 1994] between the ASK\(_{p}\) and HUI3 emotion and speech would confirm convergent and divergent validity respectively.

**Exploration of Ceiling and Floor Effects**

These same ASK, HUI3 and CHAQ data were later entered into an exploratory analysis to look for ceiling effects and floor effects. Such an exploration assumes knowledge of which children have “true” disability. Clinicians’ global ratings were previously stipulated to be the standard for the assessment of criterion validity. Although these ratings may lack precision regarding the degree of disability, it was not expected that clinicians would classify children with disabilities as non-disabled. Thus, global ratings were also used as the primary standard for this assessment.
Ceiling effects were expressed for each of the scales (ASK, CHAQ, and combined HUI3 ambulation, dexterity and pain sub-scales) as the proportion of children who scored 100% (defined as full physical function) yet were classified as disabled by the referring clinician. Floor effects for each of the scales were expressed as the proportion of children who scored 0% (defined as no physical function) on the scale. These data were also used to determine how the ASK compared to the CHAQ and to the HUI3 (ambulation, dexterity and pain sub-scales combined) in terms of its sensitivity to small amounts of disability. This was achieved by looking at the distribution of ASK scores for children reported to be disabled at the time of referral, but classified as non-disabled by the CHAQ. This analysis was repeated for the subset reported to be non-disabled by the HUI3 (ambulation, dexterity and pain sub-scales combined). The number of children who scored less than 100% on the ASK and yet were classified as non-disabled by the HUI3, was calculated to explore whether there may be a component of hidden disability in the Ontario demographic data. Note that a formal calculation of specificity and sensitivity was not possible since the sampling strategy did not include both children with and without disabilities, but only those thought to be disabled according to a clinician’s global assessment.

**Construct Validity III (responsiveness)**

In addition to criterion validity and the various forms of construct validity already described, responsiveness was considered to be an important measurement property at the monitoring level. Since the responsiveness of a measure will vary depending on the clinical population and therapeutic intervention of interest, it was not the intent of this study to determine the exact magnitude of responsiveness of the ASK for use in sample size estimation for specific clinical trials. Rather, three pieces of information were sought. (1) What is the ASK’s responsiveness in the children studied? (2) How does the ASK’s responsiveness compare to another measure’s responsiveness? (3) How strong is the correlation between change in clinical course and ASK change scores (an aspect of construct validity).
Sample Size for Sub-Sample II

Because the intent of this portion of the research is not hypothesis testing, and since the magnitude of the change in ASK scores and the standard deviation of ASK change scores were unknown, a statistical calculation of sample size necessary to prove change was neither necessary nor possible. As previously discussed under the heading of construct validity, a sample size of 24 was expected to be sufficient to prove a correlation >0.5 between ASK and clinical course (α=0.05). The sample size for this assessment was based on the third question; the correlation between clinical course and change scores. An attempt was made to increase this sample to 30 children, to establish with greater confidence the mean and standard deviation parameters for the ASK change scores.

These children were selected from those who were expected to change in a clinically important way following the completion of their first set of questionnaires, and were not expected to participate in any other secondary component of the study (e.g., not being clinically assessed for criterion validity). A global assessment of change was collected from the referring clinician at the time of initial recruitment to facilitate the identification of eligible children, and confirmed by chart review after completion of data collection.

Data Collection

Families of eligible children were contacted by phone to confirm their continued consent, and to confirm that clinically important change had occurred since the first data collection was completed. These children were mailed a second data collection package, containing a global assessment form (asking about change and current status), the ASKp, ASKc, CHAQ, and postage paid return envelope. The data were entered and aggregated in the same way as previously described.
Analysis

Analysis began with the calculation of effect size. There is some confusion over the term “effect size” (e.g., [Cohen 1988, Kazis et al. 1989]), that has also been referred to as standardized response mean [Katz et al. 1992]. In this thesis the term effect size is defined as the average size of the difference divided by the standard deviation of the difference [page 48 Cohen 1988]. The relative responsiveness was calculated by dividing the effect size for each of the ASK versions by the effect size for the CHAQ. Finally, the analysis for confirmation of construct validity was done by Pearson’s correlation between the change scores and the predicted change (ranging from negative one to positive one).

PART B: EXPLORING THE DETERMINANTS OF PHYSICAL DISABILITY

The ASK is intended to offer an alternative outcome measure that takes into account the community environments in which children live; a unique vantage point that reflects the real world as perceived by children. The ASKp measures what children report the frequency with which they have been performing the same activities in their usual environment, while the ASKc measures what children report they can do in an implied hypothetical environment (that is often an ideal environment, but could potentially be a real environment).

Because the ASK includes only activities usually performed daily by healthy children, one would expect that if a child ‘could do’ the activity, they would report doing it ‘all of the time’. However, previous empirical comparisons of performance and capability have shown the assumption that capability equals performance is not true [Anderson et al. 1977, Young 1994]. The previous research was unable to confidently establish the exact magnitude of the discrepancy, and did not explore reasons for the discrepancy. The present research moved
forward from the previous research base and compared data regarding children’s performance and capability from the ASKc and ASKp by both quantitative and qualitative methods.

**QUANTITATIVE COMPARISON of PERFORMANCE & CAPABILITY**

The quantitative comparisons of performance and capability included comparisons of summary score distributions and item structures.

**Comparison of Summary Score Distributions**

Previous comparisons of performance and capability [Anderson et al. 1977, Young 1994] have made the comparisons based on aggregations from binary response options. Such aggregations permit the purest comparisons, because both performance and capability can best be explained in binary terms (e.g., someone does something or does not do it, versus can do or cannot do). Binary response options have been used in other health status measures (e.g., the Sickness Impact Profile and the Pediatric Evaluation of Disability Inventory). Collection of data in binary form was not possible for the ASK, because item generation interviews revealed that children do not see their physical function as a dichotomy. There is a large and potentially important gray zone in between. Pilot testing of the ASK with children confirmed this, and showed that children would not answer questions if the correct response option (which lay somewhere between do always and never) was not included for them.

Based on this information, during the initial ASK development research [Young 1994] the first of the two intended binary response options (do all of the time) was retained in the ASK-performance, but the second option (never do) was divided into four categories: do ‘most of the time’, ‘sometimes’, ‘once in a while’ and ‘none of the time’. For similar reasons the ASK-capability was also given a five-category ordinal response set. However, the second response option (cannot do) was not amenable to sub-division, thus the first option (can do) was subdivided into: could do with ‘no problem’, ‘a little problem’, ‘a medium size problem’ and with ‘a big problem’. It was these five-point ordinal formats of the ASKp and ASKc that were used
to collect data accurately, however, the data were recoded into the binary formats (i.e., ASKc binary=can do vs. cannot do, ASKp binary=do all of the time vs. never do) prior to aggregation of proportion scores that were used for the purpose of comparing performance and capability.

**Power Calculation**

Data from the ASK’s previous evaluation on 28 children [Young 1994] showed the effect size between performance and capability based on proportion scores was 0.98 (d=17.4, var=17.7). Thus, application of the sample size formula, previously presented for the assessment of discriminant validity, yielded a sample size of 17.

\[
n = \frac{2(Z\alpha + Z\beta)^2 \cdot \text{var}^2}{(\mu_{\text{ASKc}} - \mu_{\text{ASKp}})^2 \cdot \text{Effect Size}^2} = \frac{15.68}{0.98^2} = 16.3
\]

However, inclusion of a larger sample would enhance the confidence in the difference estimate. Therefore, all data available from the initial data collection session was used (n=200), and the calculation was re-run to solve for Z\beta where n was 200.

\[
Z\beta = \frac{n}{2 \times \text{Effect Size}} = \frac{200}{1.41 \times 0.98} - 1.96 = 8.25
\]

The Z\beta value was then translated into a linear value to determine power, and the power of the test was expected to exceed 0.99.

**Data Collection**

No new data were required for the comparison of performance and capability. Thus, the collection methods are those described in previous sections.

**Analysis**

Comparison of performance and capability summary scores was done using two different summary scores. First, proportion scores were used. These were the proportion of items the child reported they did all of the time (ASKp) and the proportion of items they could do (ASKc). The proportion scores took into account the number of items answered (by dropping
Performance and capability would have implications for the understanding of the differences accurately predicted from capability scores. The strength of any relationship between methods. This analysis was done to determine whether or not performance scores could be observed between proportion scores.

The difference between the ARK and AVK, although of lesser magnitude than the difference in analyses. The percentile scores were expected to show a statistically significant proportion set by Rasch analysis. Thus, assessment of these differences was included as another analysis. However, for all other purposes the ARK is a better basis for the full response and point. Thus, the comparison of proportion scores was performed on the primary and point. The comparison of proportion scores was consistent with previous research and more directly comparable from a conceptual standpoint.

The rationale for these two different aggregation methods was as follows. The proportion scores were used in the aggregations.

The point response options were used in the aggregations.

First, for response options were coded as one (capable) and the last response option was coded as zero (uncapable). Second, the first response option (all of the time) was coded as one (performed) and the other responses were coded as zero (not performed, consistently). For the AVK, the 1994[1]. For the AVK, the first response option (all of the time) was coded as one (performed) and previously [237, 72] and previously [237, 72] were reduced to binary responses prior to these aggregations. Based on our point scores with missing values or those rated not-applicable. The original raw responses (zero to
Comparison of Item Structure

In addition to comparison of children's summary scores, the comparison of item structure (i.e., distribution of logit values) between the ASKp and ASKc versions was of interest. The structure of a measure can be identified by item characteristic analysis such as Rasch analysis previously described. The item characteristics for the ASKp and ASKc were generated separately. These item structures from the ASKp and ASKc were compared in this section to determine the stability of the item locations when assessed under different environments (usual vs. hypothetical). The hypothesis was that item structure between the two versions would be similar (as defined by a correlation greater than 0.5).

Sample Size

This aspect of the research used the sample previously described. Since 150 cases were expected to be more than adequate for a stable Rasch model [Linacre 1994, Wright 1996] and more cases would increase the confidence in the item structure described, all 200 children's data were used. This number was also more than sufficient to permit detection of a correlation statistically greater than 0.5 between models, if the underlying correlation was at least 0.55 [Cohen 1988].

Data Collection

No new data were required for this analysis, thus the collection methods are those described in previous sections.

Analysis

Item structure was compared between the two ASK versions using the raw data (five response options). Not-applicable responses were coded as missing data for this analysis. Item logit values were generated using the Bigsteps computer program, partial credit model, as
described in the item characteristic analysis section (methods section). The item logits were compared between the ASKa and ASKc using a paired t test. The use of parametric analysis is facilitated by the normal distribution of logits created by the Rasch model. Although the items contained in the two versions were the same items, the framing of the questions was different. Previous sections of this thesis were designed to determine how great of an effect the context difference had on scores. Similarly, this section was designed to determine how great an effect context had on the modeling of item difficulty. A significance value (p value) close to one would confirm the similarity of item structure across environments. A low significance value would require further exploration.

**QUALITATIVE COMPARISON of PERFORMANCE & CAPABILITY**

Further exploration of the difference between performance and capability was addressed by qualitative methods. Qualitative methods have traditionally been used to study social phenomena [Bowers 1988] and to generate new middle range substantive theory [Corbin and Strauss 1990]. These methods are increasingly being applied to study medical phenomena (e.g., [Britten 1995, Jones and Hunter 1995, Keen and Packwood 1995, Kitzinger 1995, Mays and Pope 1995a, Mays and Pope 1995b, Pope and Mays 1995]) for the same purpose. (The following references provide background on the basics of qualitative research: [Charmaz 1990, Conrad 1990, Gerhardt 1990, Strauss and Corbin 1990]). The qualitative component of this thesis was designed to assist in the appropriate interpretation of the quantitative comparisons of performance and capability, and in doing so, delineate factors associated with the expression of childhood physical disability.

Since there can be no discrepancy between performance and capability in children who have no capability to do an activity, any discrepancy must be associated with activities for which the children had some capability but performed the activity less than daily. Among children with unstable levels of symptoms (e.g., children with arthritis who have good days and bad days), the discrepancies may be due to fluctuations in the symptoms of disease. In
children with stable disorders (e.g., cerebral palsy) the discrepancy may be associated with a change in situation that caused their impairments (e.g., difficulty walking) to translate into disability in some environments but not in others. Important information regarding the relationship between performance and capability was sought from a sub-group of children with stable and unstable disorders.

Sample Size for Sub-Sample III

Sample sizes for qualitative research are not calculated in the same way as for quantitative comparisons. Sample sizes for qualitative analysis commonly range between five and 40 subjects (e.g., [Disman 1985, Law 1992, Locker and Kaufert 1988, Weitz 1989, Yoshida 1991]). Based on the number of interviews required for item generation (n=20) [Young 1994], it was estimated that approximately 20 interviews would be necessary. Therefore the upper limit was set conservatively at 40 children. The determination of specific sample size was based on saturation [Bowers 1988, Conrad 1990, Strauss and Corbin 1990]. When no new themes or concepts were being generated from sequential interviews, no additional interviews would be conducted.

The interviews were conducted with children who had previously completed the first data collection session in this study, and were not scheduled to participate in any additional sub-components of the study. While, they represented a sample of convenience, theoretical sampling [Charmaz 1990, Gerhardt 1990, Strauss and Corbin 1990] was employed. Attention was paid to the demographics of each subject to ensure that a range of ages, disability levels and diagnostic groups was represented. Because of the necessity to do face to face interviews, these children were recruited from within 100 km of Toronto.

Data Collection

The children were interviewed in their homes (or alternative locations), because previous research regarding disability in this population found benefit from this mechanism [Young 1994]. It was thought that the convenience to the families would improve willingness to spend
considerable time discussing difficult issues, and that environmental cues may facilitate recall of factors contributing to disability.

The children were asked a series of open-ended questions designed to help them identify areas of discrepancy between capability and performance, by describing the situations in which their performance had varied in the past. The analysis of the situation focused on sources of variation and was used to infer factors associated with the discrepancy between performance and capability across activities and across situations. In addition to the open-ended questions, children were shown their ASKp and ASKc questionnaires, and asked to talk about specific activities on which they had reported a discrepancy. All interviews were audio-taped. The tapes were transcribed verbatim by a medical records transcriptionist and verified by the principal investigator prior to being used for qualitative analyses.

Analysis

The interview transcripts were imported into the First Order Perspective (FOP) qualitative analysis program [Booth 1993]. The transcripts were segmented into small but meaningful segments of text, two to three sentences long. Each segment was labeled according to the concepts described by the text. A segment of text could have multiple labels attached to represent multiple concepts in the text. After several interviews had been completed, segmented and labeled, the labels were reviewed for similarities. Consistent codes were established by the principal investigator based on this review, and these codes were used to replace the preliminary labels. For example, segments initially labeled as having to do with ‘wet ground’ and ‘loose footing’ may have later been given a consistent code of ‘terrain’.

To protect the children’s perspectives and prevent the researcher from influencing the outcome of the research, several steps were taken. These included: reserving the detailed review of the literature until after the qualitative model had been constructed from the children’s perspectives [Strauss and Corbin 1990], and having a second researcher (KKY) independently confirm the coding of categories and analyses to ensure stability of the coding [Weber 1985].
To ensure consistency, the physical process of coding was done by the principle investigator. The coded segments were then used in a content analysis based on Grounded Theory Method (GTM) approach [Glaser and Strauss 1967]. Using GTM’s, the researcher generates core variables based on categories that account for most of the variation in the pattern of children’s statements about activities [Strauss et al. 1984]. This was achieved reviewing all text segments with the same codes. Text segments that shared similarities were considered to be related. A pencil and paper flow chart or map was used to document the relationships between codes containing similarities. In this way core variables were used to determine themes (defined as the conceptual linkages between core variables) [Strauss et al. 1984] among explanations given by children for differences between performance and capability.

The process of interviewing, coding, and thematic analysis occurred iteratively. Specifically, after each interview the tapes were transcribed, analyzed and entered into the flow-chart. Themes from the flow chart that were unclear were taken forward into the next interview, where specific activities were discussed in hopes of increasing the clarity of a code or theme. This thematic, or grounded theory analysis [Strauss and Corbin 1990] looked initially at all individuals and all activities as a group, and later attempted to distinguish whether these strategies were consistent within one individual across different activities. This analysis was hypothesis generating rather than conclusive, and was expected to enhance the contribution of children’s perceptions of disability to the present research.

**SUMMARY OF METHODS**

This chapter has presented the methods to determine the measurement properties of the ASK (objective #1) by assessing its item characteristics and various aspects of validity, including responsiveness. It has also presented the methods for determining the relationship between capability and performance (objective #2) by both quantitative and qualitative methods. The various sub-studies were summarized in Table 4 on the second page of this chapter and are depicted in Figure 1 on page 81.
Figure 1 Methods Flow Chart

Populations of 5 to 15 year old with musculoskeletal limitations at the Bloorview MacMillan Centre, Erin Oak, the Hospital for Sick Children, and Variety Village.

~ 115 mild + ~ 115 moderate + ~ 56 severe

~ 286 Children Recruited
(assuming ≥ 70% response rate)

Sub-study Number  Sub-study Description
#1 Item Characteristic Analysis based on Rasch Modeling
#3 Criterion Validity II comparison to clinicians' global ratings of disability
#4 Construct Validity I comparison to CHAQ
#5 Construct Validity II comparison to HUI3 sub-scales
#7 Quantitative Comparison of Determinants of Disability statistical comparison of \( \text{ASK}_p \) & \( \text{ASK}_c \) scores
#2 Criterion Validity I comparison of child-report to clinician-reported \( \text{ASK}_c \) based on observation
#6 Construct Validity III comparison to clinical course determination of effect size and effect size relative to CHAQ
#8 Qualitative Exploration of Determinants of Disability analysis of interview transcripts

Note:
Sub-studies #1 to #6 relate to Objective 1 (measurement properties of the ASK).
Sub-studies #7 to #8 related to Objective 2 (capability-performance difference).
CHAPTER 4

RESULTS PART A

The results of the present research are presented in the same order as the methods; beginning with the item characteristic analysis, then validity (including responsiveness), quantitative comparison of capability and performance, and finally qualitative exploration of factors responsible for the difference between capability and performance. The results pertaining to the measurement properties of the ASK are presented in this chapter. The results pertaining to the exploring the determinants of physical disability are presented in Chapter 5.

ASK MEASUREMENT PROPERTIES

The assessment of the ASK’s measurement properties required that 286 children between the ages of five and 15 years be recruited, with the expectation that 200 children would complete the baseline data collection. From this total sample, it was expected that one sub-sample of 24 children would also participate in a clinical assessment to ascertain criterion validity. A second distinct sub-sample of 30 children was expected to complete a second data collection after a clinically important change to assess responsiveness. A third distinct sub-sample was expected to participate in interviews exploring the factors that distinguish capability from performance (this latter sample is discussed in Chapter 5). These sample sizes were achieved. A breakdown of the numbers recruited and completed is shown in Table 6.

In total 200, or 71% of the 281 eligible children completed the baseline data collection. From these 200, approximately 40 children were requested to attend a clinical interview, of whom 24 (60%) consented and attended. A different group of 56 children were requested to complete a second data collection after clinically important change and 37 (66%) completed
this. However, three of these were not included in the analysis due to an excessive amount of missing data, resulting in a final responsiveness sample of 34 children (60%).

Table 6 Breakdown of Recruitment & Completion

<table>
<thead>
<tr>
<th>Description</th>
<th>Running Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>331 Children Identified</td>
<td>331 Referred</td>
</tr>
<tr>
<td>10 Ineligible† (6 cognitive, 2 language, 2 non-disabled)</td>
<td></td>
</tr>
<tr>
<td>14 Unable to contact for consent</td>
<td>307 Potentially Eligible</td>
</tr>
<tr>
<td>26 Recovered from disability before completion</td>
<td>281 Eligible</td>
</tr>
<tr>
<td>15 Refused consent (3 mild, 3 moderate, 9 unknown)</td>
<td></td>
</tr>
<tr>
<td>11 Withdrew from the study (4 mild, 4 moderate, 3 unknown)</td>
<td></td>
</tr>
<tr>
<td>4 Data lost in return mail (1 mild, 2 severe, 1 unknown)</td>
<td></td>
</tr>
<tr>
<td>2 Invited for interviews only (2 severe)</td>
<td></td>
</tr>
<tr>
<td>49 Did not complete (18 mild, 15 moderate, 2 severe, 14 unknown)</td>
<td></td>
</tr>
<tr>
<td>200 Completed Questionnaires</td>
<td>200 Completed</td>
</tr>
</tbody>
</table>

† Determined by author based on information provided by the referring clinician and screening questions posed to parents at the time of consent.

DESCRIPTIVE STATISTICS

The children in the present research ranged in age from five to 15 years, and were experiencing physical disabilities as a result of a variety of clinical conditions. The group included children with neuro-muscular conditions (44.0%), orthopaedic conditions (35.3%), and arthritic conditions (20.7%). Some of the children had acute conditions, such as fractures (8.6%) and Guillain Barre (1.5%), but more of the children (approximately 66%) had chronic conditions such as amputations (2.0%) and spina bifida (25.8%). Although the range of diagnoses was wide, these children were similar in that they were all experiencing limitations in activity because of physical impairments, and all were able to participate in answering the questionnaires.

The primary dependent variable in this thesis was the ASK performance (ASKp). Also included as an outcome was the ASK capability (ASKc). The measurement properties of these two measures were assessed and compared to clinicians’ global ratings of severity, the CHAQ,
and the HUI3. The descriptions of the children in the total sample and each of the two sub-samples for this segment of the research are shown in Table 7, with respect to these variables.

For categorical variables (e.g., gender and severity) frequencies and percentages are shown. For continuous and ordinal variables (e.g., age and summary scores) means, standard deviations, ranges and medians are shown.

Table 7 Distribution of Sample Characteristics

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Total</th>
<th>Criterion Validity</th>
<th>Responsiveness Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 girls</td>
<td>100%</td>
<td>12 girls</td>
<td>16 girls</td>
</tr>
<tr>
<td>Severity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 non-disabled</td>
<td>0.5%</td>
<td>0 non-disabled</td>
<td>0 non-disabled</td>
</tr>
<tr>
<td>71 mild</td>
<td>36%</td>
<td>10 mild</td>
<td>8 mild</td>
</tr>
<tr>
<td>85 moderate</td>
<td>43%</td>
<td>11 moderate</td>
<td>12 moderate</td>
</tr>
<tr>
<td>13 severe</td>
<td>7%</td>
<td>3 severe</td>
<td>3 severe</td>
</tr>
<tr>
<td>30 unclassified</td>
<td>15%</td>
<td>0 unclassified</td>
<td>11 unclassified</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max. range 5-15.9 yrs.</td>
<td>range 5-15.9 range 5-15.9</td>
<td>range 5-15.1 range 5-15.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.1 (SD=3.09)</td>
<td>10.2 (SD=3.33)</td>
<td>9.6 (SD=2.77)</td>
</tr>
<tr>
<td>ASKp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max. range 0-100</td>
<td>range 5-15.9 range 5-15.9</td>
<td>range 5-15.1 range 5-15.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>68.2 (SD=24.12)</td>
<td>68.8 (SD=24.05)</td>
<td>66.4 (SD=25.58)</td>
</tr>
<tr>
<td>ASKc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max. range 0-100</td>
<td>range 5-15.9 range 5-15.9</td>
<td>range 5-15.1 range 5-15.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>76.8 (SD=20.71)</td>
<td>75.6 (SD=23.19)</td>
<td>75.6 (SD=22.63)</td>
</tr>
<tr>
<td>CHAQ*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max. range 0-100</td>
<td>range 5-15.9 range 5-15.9</td>
<td>range 5-15.1 range 5-15.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>52.0 (SD=29.17)</td>
<td>52.2 (SD=30.76)</td>
<td>57.2 (SD=30.80)</td>
</tr>
<tr>
<td>HUI3 Ambulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max. range 1-6</td>
<td>range 5-15.9 range 5-15.9</td>
<td>range 5-15.1 range 5-15.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.0 (SD=2.04)</td>
<td>3.5 (SD=2.04)</td>
<td>2.4 (SD=1.95)</td>
</tr>
<tr>
<td>HUI3 # of Domains</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max. range 0 to 8</td>
<td>range 5-15.9 range 5-15.9</td>
<td>range 5-15.1 range 5-15.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0 (SD=1.36)</td>
<td>2.3 (SD=1.50)</td>
<td>1.20 (SD=0.80)</td>
</tr>
</tbody>
</table>

* transformed to 0 to 100 scale, where 100 indicates non-disabled

ASK = Activities Scale for Kids: performance and capability versions
CHAQ = Childhood Health Assessment Questionnaire
HUI3 = Health Utilities Index Mark III

Analyses of normality for each of the main variables in this thesis (the ASKp, ASKc, global rating of severity, CHAQ and HUI3) were performed on the Solo Statistical System©. The Martinez-Inglewicz test of normality (based on the median and dispersion, where

normality is indicated by $I=1.0$ and the D’Agostino Omnibus test of normality (based on skewness and kurtosis, where normality is indicated by $K^2=0.0$) were used. The Martinez-Inglewicz test recommended accepting normality for severity ($I=0.98$), ASKp ($I=0.96$), CHAQ ($I=0.93$), HUI3 ambulation sub-domain ($I=0.93$), and HUI3 number of domains ($I=1.04$), and rejecting normality for the ASKc ($I=1.06$). However, the D’Agostino Omnibus test recommended rejecting normality for all (severity $K^2=6.0$, ASKp $K^2=38.8$, ASKc $K^2=20.7$, CHAQ $K^2=60.6$, HUI3 ambulation $K^2=1660.2$, HUI3 number of domains $K^2=17.3$). Thus, the distributions of these key variables were equivocal. The distribution of ASKp, ASKc and CHAQ scores are shown graphically in Figure 2.

Figure 2  Distribution of Disability Scores

Given the equivocal results from the normality testing, the robustness of parametric methods was reviewed in more detail. The techniques used in the present research are based on the central limit theory [Altman 1991]. Since multiple samples from the population are expected to yield a normal distribution, the methods protect against failures of normality provided the sample sizes are equal, sufficiently large, and the variances are similar [Tabachnick and Fidell 1983]. Since all analyses used paired data from 200 children, the first
two specifications have been met. Table 7 shows that for variables on the same scale (i.e., the ASKp ASKc and CHAQ) the variances are roughly similar. Thus, parametric analyses (and the t test in particular [Boneau 1960]) were expected to be sufficiently robust to accommodate the mild to moderate deviations from normality experienced in this thesis’ data.

This assumption of adequate robustness was tested in this data set by applying both non-parametric and parametric methods. Although some results differed slightly, the differences were at the second and third decimal place, thus cause no differences in the conclusions. Therefore, based on the literature and empirical comparison between parametric and non-parametric results in this data set, parametric methods were used throughout this thesis, and non-parametric equivalents are reported where there was meaningful difference between the parametric and non-parametric methods.

The distributions of children in the sub-samples were compared to those of the children from the total sample using chi square for categorical data (severity and gender) and t tests for continuous data. No statistically significant differences were observed between the criterion validity assessment sample and the total sample (all p values ≥ 0.25) with the exception of the number of HUI3 domains for the responsiveness sample being significantly less than for the total sample (p=0.01). Thus, these two sub-samples were considered representative of the total sample, while acknowledging that the children in the responsiveness sub-sample were slightly less disabled than the other children (see Table 7).

ITEM CHARACTERISTIC ANALYSIS

The first assessment of the ASK’s measurement properties focused on the properties of the individual items. The intent was to determine if there were problematic items. This was achieved by Rasch analysis of the baseline data set (n=200) to determine each item’s difficulty level and degree of fit to the expected model of response patterns.
Item Location Plots & Item Characteristic Curves

Rasch analysis, run on the raw data from the baseline data collection and assuming that each item had its own response set (termed ungrouped model), demonstrated that the probability of children rating most items as one, two or three was low for both the ASKp and ASKc. The distribution of responses by response category are shown in Table 8.

Table 8  Item Response Frequencies for ASKp From 200 Children

<table>
<thead>
<tr>
<th>Item Number and Description</th>
<th>Frequency of Response (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Item Number and Description</td>
<td>none of the time</td>
</tr>
<tr>
<td>1 turned on the tap</td>
<td>10.0</td>
</tr>
<tr>
<td>2 combed my hair</td>
<td>10.0</td>
</tr>
<tr>
<td>3 prepared and brushed my teeth</td>
<td>7.0</td>
</tr>
<tr>
<td>4 got on &amp; off the toilet</td>
<td>21.0</td>
</tr>
<tr>
<td>5 used the toilet paper &amp; flushed toilet</td>
<td>17.0</td>
</tr>
<tr>
<td>6 got in &amp; out of the tub or shower</td>
<td>32.0</td>
</tr>
<tr>
<td>7 washed my whole body</td>
<td>15.0</td>
</tr>
<tr>
<td>8 got my clothes out</td>
<td>25.0</td>
</tr>
<tr>
<td>9 put my shirt on</td>
<td>7.0</td>
</tr>
<tr>
<td>10 put my pants on</td>
<td>17.0</td>
</tr>
<tr>
<td>11 fastened my clothes (buttons &amp; zippers)</td>
<td>12.0</td>
</tr>
<tr>
<td>12 put socks on my</td>
<td>15.0</td>
</tr>
<tr>
<td>13 put my shoes on and did them up</td>
<td>23.0</td>
</tr>
<tr>
<td>14 got in &amp; out of an automobile</td>
<td>31.0</td>
</tr>
<tr>
<td>15 got on &amp; off a bus</td>
<td>19.0</td>
</tr>
<tr>
<td>16 went on trips or vacations</td>
<td>4.0</td>
</tr>
<tr>
<td>17 ate my meals (used a knife &amp; fork)</td>
<td>2.0</td>
</tr>
<tr>
<td>18 used my finger to eat small foods</td>
<td>2.0</td>
</tr>
<tr>
<td>19 made a snack</td>
<td>26.0</td>
</tr>
<tr>
<td>20 poured myself a drink</td>
<td>21.0</td>
</tr>
<tr>
<td>21 took care of my medical needs</td>
<td>17.0</td>
</tr>
<tr>
<td>22 walked without any support</td>
<td>35.0</td>
</tr>
<tr>
<td>23 got around inside my home</td>
<td>5.0</td>
</tr>
<tr>
<td>24 walked/rolled in crowds</td>
<td>7.0</td>
</tr>
</tbody>
</table>

MV = Missing Value   NA = Not-Applicable

continued...
Table 8 continued... (Item Response Frequencies for ASKp From 200 Children)

<table>
<thead>
<tr>
<th>Item Number and Description</th>
<th>Response Score and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>25 walked/rolled indoors &amp; kept up with friends</td>
<td>12.0</td>
</tr>
<tr>
<td>26 got around outside</td>
<td>19.0</td>
</tr>
<tr>
<td>27 walked/rolled up a hill</td>
<td>8.5</td>
</tr>
<tr>
<td>28 walked/rolled on rough or slippery surfaces</td>
<td>13.0</td>
</tr>
<tr>
<td>29 ran/rolled outdoors &amp; kept up with friends</td>
<td>16.0</td>
</tr>
<tr>
<td>30 walked/rolled what distance</td>
<td>8.0</td>
</tr>
<tr>
<td>31 carried a drink or food to the table</td>
<td>15.0</td>
</tr>
<tr>
<td>32 carried things in 2 hands</td>
<td>24.0</td>
</tr>
<tr>
<td>33 carried my backpack</td>
<td>10.5</td>
</tr>
<tr>
<td>34 climbed up &amp; down one step or curb</td>
<td>22.5</td>
</tr>
<tr>
<td>35 climbed up &amp; down a flight of stairs</td>
<td>26.0</td>
</tr>
<tr>
<td>36 climbed up &amp; down 2 flights of stairs</td>
<td>24.5</td>
</tr>
<tr>
<td>37 played quiet activities</td>
<td>3.5</td>
</tr>
<tr>
<td>38 ran and played outside with friends</td>
<td>24.0</td>
</tr>
<tr>
<td>39 played sports I enjoy by myself</td>
<td>32.5</td>
</tr>
<tr>
<td>40 participated in gym class</td>
<td>21.0</td>
</tr>
<tr>
<td>41 played sports and active games</td>
<td>31.0</td>
</tr>
<tr>
<td>42 did my usual job or chores</td>
<td>11.0</td>
</tr>
<tr>
<td>43 got in &amp; out of a chair</td>
<td>14.0</td>
</tr>
<tr>
<td>44 sat on the floor</td>
<td>9.0</td>
</tr>
<tr>
<td>45 sat for at least 1 hour</td>
<td>3.5</td>
</tr>
<tr>
<td>46 got in &amp; out of bed</td>
<td>16.5</td>
</tr>
<tr>
<td>47 got down onto the floor &amp; up again</td>
<td>27.0</td>
</tr>
<tr>
<td>48 had fun on the floor or ground</td>
<td>17.5</td>
</tr>
<tr>
<td>49 stood still for 10 minutes</td>
<td>34.0</td>
</tr>
<tr>
<td>50 stretched to reach a high shelf</td>
<td>10.5</td>
</tr>
<tr>
<td>51 picked up things from the floor</td>
<td>10.0</td>
</tr>
<tr>
<td>52 turned door knobs</td>
<td>9.0</td>
</tr>
<tr>
<td>53 got through heavy doors</td>
<td>23.5</td>
</tr>
</tbody>
</table>

MV = Missing Value   NA = Not-Applicable

Table 8 shows that on average 3.06% of responses were missing on the ASKp and 8.11% of responses were reported as not-applicable. For comparison, the average frequency of missing values for the ASKc items was 3.97% (not-applicable responses were not allowed on
the ASKc), and the average frequency of missing values and not-applicable responses for the CHAQ were 4.66% and 4.71% respectively. Because the rate of missing and not-applicable responses combined was high, the effect of these values on subsequent analysis was assessed. This was done by calculating summary scores based on only those items with fewer than 20% blank cells (defined as missing and not-applicable values combined) and repeated with only those items with fewer than 10% blank cells.

Table 8 shows that there were 44 items with fewer than 20% blank cells, and 36 items with fewer than 10% blank cells. The summary scores calculated after excluding these items (referred to as the ASK44 and ASK36) were compared to the full 53 item summary score in which blank cells had been excluded from the numerator and denominator of the calculation. Both alternate versions of the ASKp and ASKc had Pearson correlation coefficients with their respective longer versions exceeding 0.99 (lower limit of confidence interval >0.99). This high correlation between various ASKp scores documents that the blank cells did not introduce a systematic bias into the ASKp53 summary scores. The high correlation between various ASKc scores documents that the items frequently reported as not-applicable or missing were not systematically different from the remaining items. In essence, the analysis could have been completed with either version and the same results would have been obtained. However, many of the frequently not-applicable items were important to the children [Young 1994]. Furthermore, these items were considered important to the face validity of the scale and to the conceptualization of performance as disability in a child's usual environment. Thus, the remaining analyses were based on the ASK53.

The distribution of children according to the number of items included in their summary scores is shown in Table 9. This table shows that 52.5% of the 200 children had ASKp scores calculated from raw data with less than 10% blank cells, and 87.0% had ASKp scores calculated from raw data with less than 20% blank cells. These numbers improve substantially
for the ASK44 and ASK36 alternative versions, but the change is not associated with changes in the summary scores. Thus, the proportion of blank cells is not a source of systematic bias.

Table 9 Number of Items in Children’s ASKp Summary Scores

<table>
<thead>
<tr>
<th>Number of Items in ASKp</th>
<th>Number of Children</th>
<th>Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>2</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>52</td>
<td>13</td>
<td>6.5</td>
<td>7.5</td>
</tr>
<tr>
<td>51</td>
<td>22</td>
<td>11.0</td>
<td>18.5</td>
</tr>
<tr>
<td>50</td>
<td>22</td>
<td>11.0</td>
<td>29.5</td>
</tr>
<tr>
<td>49</td>
<td>24</td>
<td>12.0</td>
<td>41.5</td>
</tr>
<tr>
<td>48</td>
<td>22</td>
<td>11.0</td>
<td>52.5</td>
</tr>
<tr>
<td>47</td>
<td>19</td>
<td>9.5</td>
<td>62.0</td>
</tr>
<tr>
<td>46</td>
<td>19</td>
<td>9.5</td>
<td>71.5</td>
</tr>
<tr>
<td>45</td>
<td>13</td>
<td>6.5</td>
<td>78.0</td>
</tr>
<tr>
<td>44</td>
<td>13</td>
<td>6.5</td>
<td>84.5</td>
</tr>
<tr>
<td>43</td>
<td>5</td>
<td>2.5</td>
<td>87.0</td>
</tr>
<tr>
<td>42</td>
<td>8</td>
<td>4.0</td>
<td>91.0</td>
</tr>
<tr>
<td>41</td>
<td>2</td>
<td>1.0</td>
<td>92.0</td>
</tr>
<tr>
<td>40</td>
<td>3</td>
<td>1.5</td>
<td>93.5</td>
</tr>
<tr>
<td>39</td>
<td>4</td>
<td>2.0</td>
<td>95.5</td>
</tr>
<tr>
<td>38</td>
<td>2</td>
<td>1.0</td>
<td>96.5</td>
</tr>
<tr>
<td>37</td>
<td>4</td>
<td>2.0</td>
<td>98.5</td>
</tr>
<tr>
<td>36</td>
<td>1</td>
<td>0.5</td>
<td>99.0</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>0.5</td>
<td>99.5</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>0.5</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

In the Rasch analysis the frequency with which each of the response options was used is important. If a response is rarely used, the model will be unable to predict it as the most likely response for any individual, and thus it contributes little information (but may contribute noise) to the Rasch model. Figure 3 shows the distribution of response option frequencies averaged across all 53 items. The blank cells are excluded from Figure 3 because they neither contribute nor detract from the Rasch model. Blank cells are accommodated in the Rasch model without
the elimination of cases and without the elimination of items. Thus, no additional data are discarded in accommodating blank cells [Fisher et al. 1995]. This is possible because the Rasch model is an iterative approach in which the difficulty level of items are first estimated as the average score for all persons answering each item. Those that did not answer the item do not contribute to the average. Then the average person ability scores are estimated based on the average of all item scores for each person. Those items that were not answered do not contribute to the person ability scores. The process is then continued in an iterative manner taking into account the estimates just generated. The model is generating probability estimates of item difficulty and person ability scores simultaneously. The more data points in the prediction, the more likely that the estimate is accurate, however, estimates do use all available information. Thus, the model maximizes its use of available data and missing data is not problematic. This is a unique advantage of the Rasch model over other item analysis methods [Fisher et al. 1995].

Figure 3 Distribution of Response Option Frequencies

Because of the infrequent use of response categories one to three across all items, the initial Rasch model was unable to predict any of these responses to be most probable, and demonstrated redundancy in the central response categories. Furthermore, while response options zero and four (unable and non-disabled) had conceptual importance, the distinctions between one, two and three were less important. Therefore, the model was repeated with the raw data recoded into three response options (0→0, 1→1, 2→1, 3→1, 4→2).
The use of response options across items was then compared, to determine whether it was reasonable to simplify the analyses and use a grouped model, in which all items are assumed to share the same response set (true) and in which the distance between response options is consistent across items. Differences were observed in the distances between response options between items. However, comparison of the item parameter estimates from grouped and ungrouped models showed no statistically significant differences (paired t test p>0.97). The remainder of the Rasch analysis was done assuming that all items shared a consistent three-option response set [Linacre 1996]. Note, however, that while their use of three responses has distinct advantages for Rasch analysis, discrimination and responsiveness may be adversely affected. Therefore, the smaller number of response options was used only to determine the item ordering and characteristics. Average summary scores, based on all five response options, were used for all other analyses.

Table 10 shows a summary of the results of item characteristics resulting from the application of a grouped Rasch model with three response options. They are listed in order of ease, with the easiest activities being shown at the top. Appendix X contains a reprint of this table with the items listed in the order that they were presented in the questionnaire (i.e., item one at the top and item 53 at the bottom of the table).

Homogeneity may be ascertained by several methods. This study used the ability of the data to converge to a Rasch model and an absence of extreme misfit as an indicator of homogeneity [Hambleton 1991, Molenaar 1995]. Convergence was achieved after eight iterations when three response options were used and no excessive misfit statistics were observed. Furthermore, previous research involving interviews with children and parents as well as a focus group with clinicians was the initial basis for postulating that the items belonged within the single domain of disability. Cronbach’s alpha was used to assess the internal consistency of the items as a final test of homogeneity and found an alpha of 0.97 (using the
ASKc because it had few missing values). It follows, therefore, that the ASK is a homogeneous measure of physical disability.

**Table 10 ASKp53 Rasch Model**

<table>
<thead>
<tr>
<th>Item Description</th>
<th>Logit</th>
<th>Error</th>
<th>Infit</th>
<th>Outfit</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 used my finger to eat small foods</td>
<td>2.35</td>
<td>0.21</td>
<td>1.52</td>
<td>1.89</td>
</tr>
<tr>
<td>17 ate my meals (used a knife &amp; fork)</td>
<td>1.25</td>
<td>0.15</td>
<td>1.10</td>
<td>1.23</td>
</tr>
<tr>
<td>23 got around inside my home</td>
<td>1.21</td>
<td>0.15</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>16 went on trips or vacations</td>
<td>1.16</td>
<td>0.15</td>
<td>1.60</td>
<td>2.64</td>
</tr>
<tr>
<td>45 sat for at least 1 hour</td>
<td>0.89</td>
<td>0.15</td>
<td>1.41</td>
<td>1.83</td>
</tr>
<tr>
<td>37 played quiet activities</td>
<td>0.79</td>
<td>0.14</td>
<td>1.08</td>
<td>1.31</td>
</tr>
<tr>
<td>9 put my shirt on</td>
<td>0.78</td>
<td>0.14</td>
<td>0.86</td>
<td>0.77</td>
</tr>
<tr>
<td>43 got in &amp; out of a chair</td>
<td>0.74</td>
<td>0.14</td>
<td>0.77</td>
<td>0.53</td>
</tr>
<tr>
<td>52 turned door knobs</td>
<td>0.63</td>
<td>0.13</td>
<td>0.80</td>
<td>0.79</td>
</tr>
<tr>
<td>1 turned on the tap</td>
<td>0.61</td>
<td>0.13</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>3 prepared and brushed my teeth</td>
<td>0.61</td>
<td>0.13</td>
<td>0.80</td>
<td>0.82</td>
</tr>
<tr>
<td>46 got in &amp; out of bed</td>
<td>0.58</td>
<td>0.13</td>
<td>0.79</td>
<td>0.57</td>
</tr>
<tr>
<td>5 used the toilet paper &amp; flushed the toilet</td>
<td>0.49</td>
<td>0.13</td>
<td>1.04</td>
<td>0.81</td>
</tr>
<tr>
<td>24 walked/rolled in crowds</td>
<td>0.46</td>
<td>0.14</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>4 got on &amp; off the toilet</td>
<td>0.44</td>
<td>0.13</td>
<td>0.94</td>
<td>0.68</td>
</tr>
<tr>
<td>51 picked up things from the floor</td>
<td>0.28</td>
<td>0.13</td>
<td>0.70</td>
<td>0.66</td>
</tr>
<tr>
<td>33 carried my backpack</td>
<td>0.27</td>
<td>0.15</td>
<td>1.14</td>
<td>1.02</td>
</tr>
<tr>
<td>44 sat on the floor</td>
<td>0.19</td>
<td>0.13</td>
<td>1.02</td>
<td>1.21</td>
</tr>
<tr>
<td>7 washed my whole body</td>
<td>0.17</td>
<td>0.12</td>
<td>0.93</td>
<td>1.04</td>
</tr>
<tr>
<td>27 walked/rolled up a hill</td>
<td>0.17</td>
<td>0.15</td>
<td>0.95</td>
<td>0.80</td>
</tr>
<tr>
<td>34 climbed up &amp; down one step or curb</td>
<td>0.11</td>
<td>0.12</td>
<td>0.90</td>
<td>0.75</td>
</tr>
<tr>
<td>10 put my pants on</td>
<td>0.09</td>
<td>0.12</td>
<td>0.71</td>
<td>0.63</td>
</tr>
<tr>
<td>11 fastened my clothes (buttons &amp; zippers)</td>
<td>0.06</td>
<td>0.14</td>
<td>0.91</td>
<td>0.85</td>
</tr>
<tr>
<td>2 combed my hair</td>
<td>0.00</td>
<td>0.13</td>
<td>1.06</td>
<td>1.43</td>
</tr>
<tr>
<td>25 walked/rolled indoors &amp; kept up with friends</td>
<td>-0.01</td>
<td>0.13</td>
<td>1.03</td>
<td>1.32</td>
</tr>
<tr>
<td>47 got down onto the floor &amp; up again</td>
<td>-0.04</td>
<td>0.12</td>
<td>0.99</td>
<td>0.82</td>
</tr>
<tr>
<td>26 got around outside</td>
<td>-0.05</td>
<td>0.13</td>
<td>1.12</td>
<td>0.98</td>
</tr>
<tr>
<td>48 had fun on the floor or ground</td>
<td>-0.06</td>
<td>0.12</td>
<td>1.56</td>
<td>1.61</td>
</tr>
<tr>
<td>12 put socks on my</td>
<td>-0.08</td>
<td>0.13</td>
<td>0.75</td>
<td>0.67</td>
</tr>
<tr>
<td>30 walked/rolled about the same distance as</td>
<td>-0.11</td>
<td>0.13</td>
<td>1.00</td>
<td>1.12</td>
</tr>
<tr>
<td>28 walked/rolled on rough or slippery surfaces</td>
<td>-0.15</td>
<td>0.14</td>
<td>1.00</td>
<td>0.90</td>
</tr>
</tbody>
</table>

continued...
<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Logit</th>
<th>Error</th>
<th>Infit</th>
<th>Outfit</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>carried a drink or food to the table</td>
<td>-0.23</td>
<td>0.12</td>
<td>0.80</td>
<td>0.81</td>
</tr>
<tr>
<td>35</td>
<td>climbed up &amp; down a flight of stairs</td>
<td>-0.29</td>
<td>0.13</td>
<td>0.92</td>
<td>0.76</td>
</tr>
<tr>
<td>21</td>
<td>took care of my medical needs</td>
<td>-0.34</td>
<td>0.14</td>
<td>1.24</td>
<td>1.32</td>
</tr>
<tr>
<td>20</td>
<td>poured myself a drink</td>
<td>-0.39</td>
<td>0.12</td>
<td>0.77</td>
<td>0.78</td>
</tr>
<tr>
<td>13</td>
<td>put my shoes on and did them up</td>
<td>-0.47</td>
<td>0.12</td>
<td>0.98</td>
<td>0.94</td>
</tr>
<tr>
<td>15</td>
<td>got on &amp; off a bus</td>
<td>-0.51</td>
<td>0.18</td>
<td>1.10</td>
<td>0.96</td>
</tr>
<tr>
<td>22</td>
<td>walked without any support</td>
<td>-0.52</td>
<td>0.12</td>
<td>1.25</td>
<td>1.32</td>
</tr>
<tr>
<td>32</td>
<td>carried things in 2 hands</td>
<td>-0.53</td>
<td>0.12</td>
<td>0.91</td>
<td>0.84</td>
</tr>
<tr>
<td>6</td>
<td>got in &amp; out of the tub or shower</td>
<td>-0.55</td>
<td>0.12</td>
<td>0.78</td>
<td>0.69</td>
</tr>
<tr>
<td>14</td>
<td>got in &amp; out of an automobile</td>
<td>-0.55</td>
<td>0.12</td>
<td>0.75</td>
<td>0.64</td>
</tr>
<tr>
<td>50</td>
<td>stretched to reach a high shelf</td>
<td>-0.55</td>
<td>0.15</td>
<td>0.70</td>
<td>0.81</td>
</tr>
<tr>
<td>8</td>
<td>got my clothes out</td>
<td>-0.56</td>
<td>0.12</td>
<td>0.64</td>
<td>0.61</td>
</tr>
<tr>
<td>53</td>
<td>got through heavy doors</td>
<td>-0.60</td>
<td>0.12</td>
<td>0.67</td>
<td>0.66</td>
</tr>
<tr>
<td>38</td>
<td>ran and played outside with friends</td>
<td>-0.65</td>
<td>0.12</td>
<td>1.37</td>
<td>1.36</td>
</tr>
<tr>
<td>36</td>
<td>climbed up &amp; down 2 flights of stairs</td>
<td>-0.67</td>
<td>0.15</td>
<td>0.83</td>
<td>0.72</td>
</tr>
<tr>
<td>40</td>
<td>participated in gym class</td>
<td>-0.69</td>
<td>0.15</td>
<td>2.03</td>
<td>2.57</td>
</tr>
<tr>
<td>42</td>
<td>did my usual job or chores</td>
<td>-0.74</td>
<td>0.15</td>
<td>0.74</td>
<td>0.84</td>
</tr>
<tr>
<td>29</td>
<td>ran/rolled outdoors &amp; kept up with friends</td>
<td>-0.80</td>
<td>0.13</td>
<td>1.27</td>
<td>1.57</td>
</tr>
<tr>
<td>19</td>
<td>made a snack</td>
<td>-0.93</td>
<td>0.12</td>
<td>0.70</td>
<td>0.66</td>
</tr>
<tr>
<td>39</td>
<td>played sports I enjoy by myself</td>
<td>-0.93</td>
<td>0.12</td>
<td>1.38</td>
<td>1.49</td>
</tr>
<tr>
<td>49</td>
<td>stood still for 10 minutes</td>
<td>-1.06</td>
<td>0.12</td>
<td>1.04</td>
<td>1.03</td>
</tr>
<tr>
<td>41</td>
<td>played sports and active games</td>
<td>-1.26</td>
<td>0.13</td>
<td>1.51</td>
<td>1.82</td>
</tr>
</tbody>
</table>

The logit values in Table 10 represent the location of items along the disability domain. Items with large positive values are very easy, whereas those with large negative numbers are hard. Thus, the table shows that although the items cover a wide range of disabilities, there are disability levels at which several items are clustered.

The infit and outfit statistics in the table are useful in determining items that are strongly associated with other items (potentially redundant) and those items with frequent unexpected responses (noisy). Benjamin Wright, who developed the Bigsteps computer program used in this analysis, suggests the following cut points for misfits. Possible redundancy is indicated by a fit statistic less than 0.5. Excessive variability, or noise, is indicated by a fit statistic.
greater than 1.7. Values below 0.8 border on redundancy and those greater than 1.2 border on noise [Wright and Linacre 1994]. Thus, the table provides information useful in identifying components of the scale that are in need of refinement.

In summary, the item characteristic analysis has contributed to the validity of the scale by confirming homogeneity within the disability construct, has identified the hierarchy within the items (including a measure of the distance between items), and has identified a few items in need of refinement. Because refinement was not a primary purpose of the present research, yet is of interest, it has been dealt with in Appendix III.

VALIDATION STUDIES

The present research assessed criterion validity against two standards, and construct validity against three constructs.

Criterion Validity I (comparison to clinician observation)

A 30 item ASKc observational version was constructed from the 53 item ASKc version for the sole purpose of comparing child-report to clinician-report. As was explained in the methods section, this was necessary because not all items were observable in a clinical setting (e.g., playing with friends). This modified version of the ASKc contained only those 30 items that were observable in a clinical setting, and was restricted to four response options (limit of observational discrimination). The ASKc observational version was administered to a sub-set of 24 children in clinic. (These children had all previously completed the full 53 item version of the ASKp and ASKc at home.) After completing their self-report, the 24 children demonstrated the activities and were rated by two clinicians (A and B) who were blind to the children's responses and each others. The mean age of these 24 children was 10.2 years (SD=3.3) and 50% of the children were female. Additional details of this population's disability levels were previously presented in Table 7.
Table II  Pearson's Correlation Matrix Between ASK Versions

<table>
<thead>
<tr>
<th></th>
<th>ASK</th>
<th>ASKP</th>
<th>ASKc</th>
<th>ASKPc</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08</td>
<td>0.82</td>
<td>0.96</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>0.96</td>
<td>0.82</td>
<td>0.96</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>N/A</td>
<td>0.92</td>
<td>0.96</td>
<td>0.93</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table II: The ASK observational version was observed only on 20 of 33 items that were observed in a clinic setting and based on data from 24 children.

Generalizability of correlation validity to the full versions of the ASK and ASKP was assessed by ICC and found to be 0.99 (95% CI Lower Limit 0.98).

To determine the generalizability of results from the ASK-observational version to the clinical setting, between clinician was assessed by ICC and found to be 0.99 (95% CI Lower Limit 0.98), thus they were not examined further. In addition, the reliability of initial training and the post-training assessments showed minimal change (ICC=0.99, 95% CI Lower Limit 0.98). This appeared to confirm the validity of correlation coefficient (ICC) for the same comparison was calculated for correlation of clinician A as per a prior specification (was 0.92 (95% CI 0.82 to 0.97). As the interclass correlation is between child-report and clinician observation (based on the Pearson's correlation between child-report and clinician observation (based on the Pearson's correlation matrix)
Given the strong evidence for validity of the ASK summary scores, the validity data were re-analyzed on an exploratory basis, at the item level, for each of the 30 items included in the observational version of the ASK.

Table 12 Summary of Item Validity and Item Reliability

<table>
<thead>
<tr>
<th>Item #</th>
<th>Item Description</th>
<th>Item Reliability</th>
<th>Item Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Kappa (Clinician A vs. B)</td>
<td>Kappa (Child vs. Clinician)</td>
</tr>
<tr>
<td>1</td>
<td>turned on the tap</td>
<td>0.89</td>
<td>0.94</td>
</tr>
<tr>
<td>2</td>
<td>combed my hair</td>
<td>0.88</td>
<td>0.60</td>
</tr>
<tr>
<td>3</td>
<td>prepared and brushed my teeth</td>
<td>0.96</td>
<td>0.88</td>
</tr>
<tr>
<td>7</td>
<td>washed my whole body</td>
<td>0.46</td>
<td>0.14</td>
</tr>
<tr>
<td>9</td>
<td>put my shirt on</td>
<td>UC (100%)</td>
<td>UC (100%)</td>
</tr>
<tr>
<td>10</td>
<td>put my pants on</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td>11</td>
<td>fastened my clothes (buttons &amp; zippers)</td>
<td>0.85</td>
<td>0.46</td>
</tr>
<tr>
<td>12</td>
<td>put socks on my</td>
<td>0.88</td>
<td>0.89</td>
</tr>
<tr>
<td>13</td>
<td>put my shoes on and did them up</td>
<td>1.00</td>
<td>0.69</td>
</tr>
<tr>
<td>17</td>
<td>ate my meals (used a knife &amp; fork)</td>
<td>0.88</td>
<td>UC (92%)</td>
</tr>
<tr>
<td>18</td>
<td>used my finger to eat small foods</td>
<td>UC (100%)</td>
<td>UC (100%)</td>
</tr>
<tr>
<td>19</td>
<td>made a snack</td>
<td>1.00</td>
<td>0.43</td>
</tr>
<tr>
<td>20</td>
<td>poured myself a drink</td>
<td>0.64</td>
<td>0.48</td>
</tr>
<tr>
<td>22</td>
<td>walk without any support</td>
<td>1.00</td>
<td>0.93</td>
</tr>
<tr>
<td>25</td>
<td>walked indoors &amp; kept up with peers</td>
<td>0.60</td>
<td>0.08</td>
</tr>
<tr>
<td>29</td>
<td>walked outdoors &amp; kept up with peers</td>
<td>0.82</td>
<td>0.70</td>
</tr>
<tr>
<td>32</td>
<td>carried a drink or food to the table</td>
<td>0.46</td>
<td>0.26</td>
</tr>
<tr>
<td>33</td>
<td>carried things in 2 hands</td>
<td>0.97</td>
<td>0.79</td>
</tr>
<tr>
<td>34</td>
<td>carried my backpack</td>
<td>0.87</td>
<td>0.61</td>
</tr>
<tr>
<td>35</td>
<td>climbed up &amp; down one step or curb</td>
<td>1.00</td>
<td>0.76</td>
</tr>
<tr>
<td>36</td>
<td>climbed up &amp; down a flight of stairs</td>
<td>1.00</td>
<td>0.89</td>
</tr>
<tr>
<td>37</td>
<td>climbed up &amp; down 2 flights of stairs</td>
<td>0.98</td>
<td>0.89</td>
</tr>
<tr>
<td>38</td>
<td>played quiet activities</td>
<td>0.85</td>
<td>UC (88%)</td>
</tr>
<tr>
<td>40</td>
<td>played sports I enjoy by myself</td>
<td>0.95</td>
<td>-0.13</td>
</tr>
<tr>
<td>44</td>
<td>got in and out of a chair</td>
<td>0.95</td>
<td>0.93</td>
</tr>
<tr>
<td>47</td>
<td>got in and out of bed</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>48</td>
<td>got down onto the floor &amp; up again</td>
<td>0.97</td>
<td>0.90</td>
</tr>
<tr>
<td>51</td>
<td>stretched to reach a high shelf</td>
<td>0.97</td>
<td>0.55</td>
</tr>
<tr>
<td>52</td>
<td>picked up things from the floor</td>
<td>UC (94%)</td>
<td>-0.05</td>
</tr>
<tr>
<td>53</td>
<td>turned door knobs</td>
<td>UC (91%)</td>
<td>UC (82%)</td>
</tr>
</tbody>
</table>

Mean Quadratic Weighted Kappa 0.87 0.61

UC = Unable to calculate due to absence of variability; value in brackets is the percentage agreement.
Reliability testing of the relationship between two clinicians' ratings on individual items resulted in an average quadratic weighted Kappa (across 30 items) of 0.87 (range 0.46-1.00). These results are also presented in Table 12. Although there were no items with poor inter-clinician reliability, there were four items for which a Kappa could not be calculated because all children had the same response on the item. These items all exceeded 90 percent agreement. Validity testing of the relationship between clinician-ratings and child-ratings on individual items resulted in an average Kappa (across 30 items) of 0.61. It also revealed five items with low Kappa values (<0.4), indicating sub-optimal agreement, and five items that had too little variability to confidently determine their validity (all five exceeded 80 percent agreement). Detailed results are shown in Table 12.

Criterion Validity II (comparison to global ratings of disability severity)

A second method for establishing criterion validity (sometimes referred to as discriminative validity or a form of construct validity) was assessed by comparing ASKp and ASKc scores across groups of children with different disability ratings. The disability ratings were collected from the referring clinicians prior to baseline data collection, and were available for 170 of the 200 children (85%). Analysis of variance (ANOVA) was applied and demonstrated a statistically significant difference across groups for both the ASKp (p<0.0001) and ASKc (p<0.0001). Comparisons were then made between groups using unpaired t tests and demonstrated significant differences between mild vs. moderate and moderate vs. severe groups on both the ASKp and ASKc (p<0.0001). These values were consistent with those from the Kruskal-Wallis test and confirmed discriminant validity of both versions of the ASK. The non-parametric test for linear trend, that is an extension of the Kruskal-Wallis test [Altman 1991], confirmed a significant linear trend across the mild, moderate and severely disabled groups (z=-9.52, p<0.0001). The means and standard deviations of ASK scores according to clinicians' global ratings of severity are shown in Table 13.
Table 13  ASK Scores by Clinicians’ Global Ratings of Disability Severity

<table>
<thead>
<tr>
<th>Severity</th>
<th>Mild n=71</th>
<th>Moderate n=85</th>
<th>Severe n=13</th>
<th>p value across all 3 groups</th>
<th>p value between any 2 groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASKp</td>
<td>85.4 (13.39)</td>
<td>56.5 (19.33)</td>
<td>26.2 (9.07)</td>
<td>p &lt; 0.0001</td>
<td>p &lt; 0.0001</td>
</tr>
<tr>
<td>ASKc</td>
<td>90.7 (10.59)</td>
<td>69.4 (17.28)</td>
<td>36.9 (12.46)</td>
<td>p &lt; 0.0001</td>
<td>p &lt; 0.0001</td>
</tr>
</tbody>
</table>

The distribution of ASKp and ASKc scores by severity is shown in Figure 4. This figure shows that the ASKp distinguished between mild, moderate and severely disabled groups slightly better than the ASKc.

Figure 4  Distribution of ASK Scores by Severity of Disability

![Box plots showing ASK performance and capability by severity](image)

Construct Validity I (comparison to CHAQ)

The validity of the ASK was first assessed by comparisons with the CHAQ. This comparison was classified as construct validity, rather than criterion validity, because the CHAQ was not considered an accepted standard for a generic musculoskeletal population. The
Pearson's correlation between versions of the ASK was 0.92 (95% CI 0.89 to 0.94), therefore, analyses for the ASKp and ASKc were expected to yield similar results. The Pearson's correlation between the CHAQ and ASKp was 0.81 (95% CI 0.76 to 0.86). The Pearson's correlation between the CHAQ and the ASKc was 0.82 (95% CI 0.76 to 0.86; see also Figure 5). This information further confirmed the construct validity of the ASK, because it exceeded the hypothesis of $r>0.5$ set a priori. Because of the strong correlation between the ASKp and ASKc, and for the sake of clarity, subsequent results will be reported only for the ASKp. Where potentially meaningful differences exist, both ASKp and ASKc results will be presented.

**Figure 5 Relationship Between the ASKp and CHAQ**

![Relationship Between the ASKp and CHAQ](image)

**Construct Validity II (comparison to HUI3)**

Convergent and divergent validity was assessed by correlation of the ASK versions with sub-scales from the Health Utilities Index Mark III (HUI3) to which it was expected to be similar (ambulation and dexterity) and to sub-scales of the HUI3 to which it was not expected to be similar (emotion and speech). The Spearman's correlations are shown in Table 14.
Table 14  Convergent & Divergent Spearman's Correlation Coefficients

<table>
<thead>
<tr>
<th>Construct</th>
<th>HUI3 Sub-scale</th>
<th>ASKp</th>
<th>ASKc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convergent</td>
<td>Ambulation</td>
<td>0.74</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>Dexterity</td>
<td>-0.10</td>
<td>-0.12</td>
</tr>
<tr>
<td>Divergent</td>
<td>Emotion</td>
<td>-0.09</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Speech</td>
<td>-0.15</td>
<td>(0.05 to 0.23)</td>
</tr>
</tbody>
</table>

Brackets contain 95% confidence intervals [page 295 Altman 1991].

The high correlation with ambulation, and low correlations with emotion and speech sub-domains of the HUI3 confirm convergent and divergent validity. The low correlation with dexterity was unexpected, and warrants further review.

A review of the data showed that the HUI3 dexterity scale identified an unexpectedly low proportion of children (18%) who had limitations in dexterity, in contrast to the 62% whose parents reported limitations in ambulation, on the HUI3. This finding suggested three possible explanations. First, the diverse mix of diagnostic groups included many children who were unlikely to have specific dexterity problems (e.g., lower extremity fractures), yet would have depressed ASK scores because of limitations unrelated to dexterity. Second, the HUI3 screening question for dexterity problems (whether the child usually has “the full use of two hands and ten fingers”) may have been insensitive to minor dexterity problems. These minor problems may have been detected by questions on the ASK (e.g., doing up fasteners on clothing), thereby depressing the ASK scores. Third, specific dexterity activities were not identified as a component of major importance to children during the ASK’s development. As a result, the ASK may contain more questions that detect ambulation problems than dexterity problems and lead to questioning the ASK’s validity. These three factors may have contributed to the observed high correlations between the ASK and HUI3 ambulation sub-scale yet lower correlations with the HUI3 dexterity sub-scale. Because the issue of the ASK’s validity has been addressed by the two assessments of criterion validity and two other assessments of
construct validity within the present research, these findings are not interpreted as a threat to the ASK's validity.

**Ceiling and Floor Effects**

An exploratory analysis was completed to determine whether the ASK had ceiling and floor effects, and how sensitive the ASK was to small amounts of disability in comparison to the HUI3 and CHAQ.

All 200 children in the study were classified as disabled by one of the *a priori* criteria (clinician's global rating, ASK score, CHAQ score or HUI3 ambulation, dexterity or pain subscale). The global rating has previously been presented as the standard, and since all but one of the 200 children were rated as disabled by global rating, 199 of the children were expected to score below the ceiling (<100 points). Since it is difficult to imagine a child whose physical disability could never worsen, all 200 of the children were expected to score above the floor (>0 points). The number of children scoring at the ceiling or floor of the various measures is shown in Table 15. The proportion of children who scored at the ceiling or floor are presented in brackets as a percentage of the total number of children with summary scores for that measure.

**Table 15 Frequency of Ceiling and Floor Effects by Scale**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Frequency at Ceiling (score=100%)</th>
<th>Frequency at Floor (score=0%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASKp</td>
<td>8 (4%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>ASKc</td>
<td>20 (10%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>CHAQ</td>
<td>13 (7%)</td>
<td>4 (2%)</td>
</tr>
<tr>
<td>HUI3 (ambulation, dexterity and pain sub-scales combined)</td>
<td>42 (22%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Referring Clinician</td>
<td>1 (0.5%)</td>
<td>13 (7%)</td>
</tr>
</tbody>
</table>
Ceiling effects were explored by looking at the distribution of scores on each of the scales for the 42 children who were classified as disabled by the referring clinician but scored at the ceiling on the HUI3 (ambulation, dexterity or pain sub-scales combined). Since this group may represent children with 'minor' disabilities, the data presented in Table 16 may be considered an exploration of sensitivity to small disability effects.

Table 16 Ceiling Effects in a Selected Subset
(Distribution of Scores Among 42 Children with Equivocal* Disability Ratings)

<table>
<thead>
<tr>
<th></th>
<th>Mean (Std. Dev.)</th>
<th>Range</th>
<th># at the Ceiling (percentage of 42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASKp</td>
<td>81.5 (18.24)</td>
<td>27-100</td>
<td>3 (7%)</td>
</tr>
<tr>
<td>ASKc</td>
<td>89.6 (13.16)</td>
<td>43-100</td>
<td>10 (24%)</td>
</tr>
<tr>
<td>CHAQ</td>
<td>72.6 (24.92)</td>
<td>8-100</td>
<td>7 (17%)</td>
</tr>
</tbody>
</table>
| Referring Clinician | 27 mild
7 moderated
8 not classified |       | 0 (0%)                             |

* classified as having equivocal disability ratings because they scored at the ceiling on HUI3 ambulation, dexterity and pain sub-scales yet were classified as disabled by their referring clinician (a priori standard for disabled status).

These results, combined with those shown in Table 15, suggest that the ASKc, CHAQ and HUI3 all have problems with ceiling effects. Some of these effects may be due to different definitions of disability (e.g., classified as disabled on the HUI3, or CHAQ because of assistive devices such as leg braces, that may not represent disability according to the ICIDH framework). In addition, five of the children scored at the ceiling on the ASKp, ASKc and CHAQ, indicating that perhaps they were truly non-disabled. If these children are excluded, the ceiling effects for the ASKp become negligible. Further exploration of ways to ameliorate ceiling effects is warranted for the ASKc more so than for the ASKp.

The results also suggest that floor effects exist only for the CHAQ. It would not have been unexpected for the HUI3 to have floor effects, since it has been used to screen for the presence of disability [Feeny et al. 1995] that may later be quantified by other measures. However, it is problematic for the CHAQ to have floor effects. Thus, the four children who
scored 0 on the CHAQ were explored further. These four children had average \( ASKp \) scores of 29.5 (range 18 to 36) and average \( ASKc \) scores of 31.5 (range 22 to 38). All these children were between nine and 15 years of age, two had Duchenne muscular dystrophy, one had developmental delay, and the other had arthrogryposis. These diagnoses suggest that the children should possess limited physical function, but that some physical ability should be present. Thus, the CHAQ appears to have floor effects, where the \( ASKp \) and \( ASKc \) do not.

Because the HUI3 has been used to estimate the prevalence of disability in Ontario, the data presented above can be used to consider the effect more precise measures of disability may have on disability prevalence rates. Of the 42 children classified as "non-physically-disabled" by the HUI3 (represented in Table 16), all were classified as disabled by the referring clinician. However, 12 were subsequently classified as non-disabled by either the \( ASKp \), \( ASKc \) or CHAQ. Thus, it appears that the HUI3 may have missed minor disabilities in many of the remaining children (42-12=30), resulting in an underestimate of the true disability rate among the 200 sampled by approximately 15%. However, among those classified as "physically disabled" by the HUI3, five scored 100 on the \( ASKp \), an additional five scored 100 on the \( ASKc \), and a further three scored 100 on the CHAQ (three of these 13 children scored 100 on all three scales). This suggests that the HUI3 may have overestimated physical disability for as many as 13 of the 200 children, resulting in an overestimation of approximately 6.5% (net rate of 8.5% underestimation). This may or may not be considered an acceptable margin of error.

**Construct Validity III (responsiveness)**

Responsiveness was assessed to determine whether or not the \( ASK \) was able to detect clinically meaningful change, and to determine how the magnitude of the change compared to that detected by a similar measure (the CHAQ). Unlike in clinical effectiveness trials, this population was heterogeneous. As a result some children's disability was expected to worsen (\( n=11 \)) and some children's disability was expected to improve (\( n=23 \)). Because of the opposite directions of the change predicted by the referring clinicians, responsiveness data
were analyzed separately for the two groups. The mean change, standard deviation of the change, and effect size (change + standard deviation of the change) for the ASKp, ASKc and CHAQ are shown in Table 17. In addition, the responsiveness based on a difference in t scores between pre- and post-intervention has been included as an alternative indication of responsiveness. Relative responsiveness has been computed with respect to the CHAQ.

Table 17  Responsiveness Results by Prognosis

<table>
<thead>
<tr>
<th>Scale</th>
<th>Change</th>
<th>Standard Deviation of Change</th>
<th>Pre-Post Correlation</th>
<th>Absolute Responsiveness Effect Size†</th>
<th>t statistic</th>
<th>Relative Responsiveness Effect Size‡</th>
<th>t statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicted to get worse (n=11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASKp</td>
<td>-33.5</td>
<td>27.2</td>
<td>0.40</td>
<td>1.59</td>
<td>-4.07</td>
<td>0.84†</td>
<td>0.96‡</td>
</tr>
<tr>
<td>ASKc</td>
<td>-30.6</td>
<td>28.3</td>
<td>0.17</td>
<td>1.18</td>
<td>-3.59</td>
<td>0.63‡</td>
<td>0.84‡</td>
</tr>
<tr>
<td>CHAQ</td>
<td>-36.5</td>
<td>28.5</td>
<td>0.54</td>
<td>1.89</td>
<td>-4.25</td>
<td>reference criteria</td>
<td>reference criteria</td>
</tr>
<tr>
<td>Predicted to improve (n=23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASKp</td>
<td>25.7</td>
<td>21.9</td>
<td>0.57</td>
<td>1.79</td>
<td>5.62</td>
<td>1.08‡</td>
<td>1.05‡</td>
</tr>
<tr>
<td>ASKc</td>
<td>18.4</td>
<td>17.3</td>
<td>0.69</td>
<td>1.91</td>
<td>5.12</td>
<td>1.15‡</td>
<td>0.96‡</td>
</tr>
<tr>
<td>CHAQ</td>
<td>29.7</td>
<td>26.6</td>
<td>0.55</td>
<td>1.66</td>
<td>5.34</td>
<td>reference criteria</td>
<td>reference criteria</td>
</tr>
</tbody>
</table>

† adjusted (divided by \sqrt{1-r}) to account for the correlation between pre and post scores [page 49 Cohen 1988]
‡ statistical significance of differences between these effect sizes and t statistics is unknown.

The preceding table shows that all three measures are able to pick up large [Cohen 1988] differences in these two groups of children. The ASKp and ASKc appear to be slightly more responsive than the CHAQ is to improvement in function and slightly less in terms of decline in function, however, the statistical significance of these differences is unknown. Overall, the ASKp and ASKc were as responsive as the CHAQ, based on effect size. Since there appears to be some debate in the clinical epidemiology literature as to the best indicator of responsiveness, t statistics were used as a comparator to the effect size results (where effect size is the standard method used in psychometrics). The t test indicator of responsiveness
shows a similar pattern relative to the CHAQ. Thus, the ASKp and ASKc appear to be responsive to clinically important change, particularly when it is in a positive direction, regardless of the method used to calculate responsiveness.

Responsiveness statistics were also computed at the item level. Since this was intended to contribute information to the final item selection process, the responsiveness of individual items is reported in Table A2 of Appendix III. Note that in order to combine the results for all 34 children, the change scores for the 11 predicted to get worse were multiplied by negative one to correct the direction of their change scores. Had this not been done, the scores from these 11 children would have diminished the average change when combined with the group expected to improve. The individual item responsiveness data suggest that all the items are detecting change in the heterogeneous musculoskeletal population on whom responsiveness was tested. Because this sub-sample of children was slightly less disabled than the total sample, it suggests that all items are responsive in the range that clinicians expressed the greatest uncertainty about (mild to moderate severity). Thus, the results support the value of the ASK over clinical assessment alone.

Finally, to assess the degree to which the ASK was consistent with the clinical predication of change, Pearson correlations were run between the ASK version and the predicted change. The ASKp was correlated with the predicted change at $r= 0.77$ (95% CI 0.58 to 0.88), and the ASKc at $r=0.75$ (95% CI 0.55 to 0.87). These correlations are comparable to that of the CHAQ ($r=0.76$; 95% CI 0.57 to 0.87).

In summary, the various forms of validity evidence suggest that the ASKp is consistent with clinician observation and the CHAQ, shows acceptable convergent validity, and good divergent validity. The ASKp is at least equally responsive to the CHAQ, yet has overcome the problem of floor effects evident for the CHAQ, and has slightly smaller ceiling effects. Results were similar for the ASKc, however, the responsiveness was slightly less than for the ASKp.
### SUMMARY OF MEASUREMENT PROPERTY RESULTS

#### Table 18 Summary of Results Part A

<table>
<thead>
<tr>
<th>Primary Objective</th>
<th>Purpose</th>
<th>Results</th>
</tr>
</thead>
</table>
| **ASK Measurement Properties** | 1 Item Characteristic Analysis - Rasch modeling | n=200 (n=200 proposed)  
- described the order and location of items according to difficulty, and degree of misfit  
- confirmed unidimensionality |
| 2 Criterion validity I relative to clinician observation | n=24 (n=24 proposed)  
- strong correlation between child-report and clinician-report (mean r=0.92, CI 0.82-0.97) |
| 3 Criterion Validity II (Discriminative Validity) across severity groups | n=200 (n≥130 proposed)  
- statistically significant difference between children with mild, moderate and severe physical disability (ANOVA p<0.0001) |
| 4 Construct Validity relative to another disability scale | n=193 (n≥24 proposed)  
- strong correlation with the CHAQ (r=0.81, 95% CI 0.76 to 0.86) |
| 5 Convergent and Divergent Validity relative to other indicators of disability | n=198 (n≥43 proposed)  
- stronger correlations with locomotion and dexterity (mean rho=0.42) than with emotion and speech (mean rho=-0.03) |
| 6 Responsiveness to clinically important change | n=34 (n≥30 proposed)  
- good responsiveness: effect size of 1.73 for the ASKp and 1.67 for the ASKc, compared to 1.73 for the CHAQ |
EXPLORING THE DETERMINANTS OF PHYSICAL DISABILITY

Having established that the ASK has sound measurement properties, this chapter presents the second parts of the results: the comparison of capability and performance that is based on the ASK.

QUANTITATIVE COMPARISON OF ASKp and ASKc

Comparison of Summary Score Distributions

Previous research had documented that capability exceeded performance by 18% based on summary scores from binary response options [Young 1994]. Thus, the first comparison to be made was the comparison of ASKp to ASKc, from the sample of 200 children, based on binary scores (referred to here as proportion scores). The justification for the binary cut points was expressed in methods chapter and has been previously established [Young et al. 1996]. Binary scores were deemed necessary to permit more pure comparison, because both performance and capability can best be explained in binary terms (e.g., someone does something or does not do it, versus can do or cannot do). The cut points were determined a priori. For the sake of distinguishing binary based scores from the ASK percentile summary scores (maximum range zero to 100) presented in previous sections, the binary data will be presented as the proportion of activities the child performed or was capable of (range zero to 100).

Comparison of ASKp and ASKc proportion data confirmed that capability exceeded performance (paired t test p<0.0001). The mean difference between performance and
capability observed in this data set (n=200) was 35 points on a scale of 0 to 100. The standard deviation was 21, and the confidence interval about the mean was 32 to 38. (The confidence interval about the median of 36 was 30 to 40.) The relationship between capability and performance based on this data is shown in Figure 6.

Figure 6  Relationship Between Capability and Performance
(Based on Binary Response Options)

This difference was also explored on the basis of the percentile summary scores (maximum range zero to 100) based on the full five point response set for both scales. This also showed that capability exceeded performance by a statistically significant margin (paired t test p<0.0001). The mean difference for the percentile scores was 8.64 (standard deviation 9.81) with a confidence interval around the mean of 7.3 to 10.0, and a confidence interval around the median of 6.0 was 5.0 to 8.0. The relationship between capability and performance based on this data is shown in Figure 7.
Thus, the data have shown that capability was statistically greater than performance regardless of the method of aggregation used. However, the method of aggregation did alter the magnitude of the discrepancy. Exploration of this difference using basic linear regression showed that performance could be predicted reasonably well from capability scores ($r^2 = 0.84$) based on percentile summary scores, but less well based on proportion summary scores ($r^2 = 0.46$). These regression analyses also suggested that the discrepancy between performance and capability was greater ($p < 0.001$) for children who had greater disability (lower scores).

**Comparison of Item Structure**

Although scores for children were different depending on whether capability and performance was measured, it was of theoretical interest to determine whether the item difficulty logits created by Rasch analysis differed between capability and performance models.
Comparison of the models was done using parametric statistics, since logit values are linear. Paired t test showed no statistically significant difference between models (p=0.99).

**QUALITATIVE COMPARISON OF PERFORMANCE & CAPABILITY**

The final sub-study of the present research explored the components of context associated with differences in physical disability, in an attempt to explain the observed discrepancy between capability and performance summary scores. Semi-structured interviews, in which children talked about the activities that they did less than ‘all of the time,’ were used to explore the reasons for these quantitative discrepancies. Activities children reported doing ‘all of the time’ were assumed to display no situational variability of relevance to the present research. Therefore these activities would offer little or no information regarding the factors responsible for variability in performance, and were not included in the interviews. During the interviews, children were asked to give examples of when they did and did not do certain activities. Essentially this process helped them to articulate the situations in which they performed differently, and thus identify the factors that lead them to give different answers on the performance and capability versions of the ASK.

**Demographics of the Interviewed Children and their Families**

The interviews were conducted between October 1995 and February 1996. Interview participants were accrued by convenience sampling from the larger study population (n=200), until saturation of key factors related to the capability-performance difference was reached. This occurred after the 21st child had been interviewed. Note that the last four children interviewed were members of one family, and only the first child in this group (interview participant #21) participated in the larger study.

These children represented an equal gender distribution, the full age range, and a wide variety of musculoskeletal diagnostic groups, as is shown in Table 19. The boys (mean age 11.3, range 5.9 - 15.9 years) were slightly older than the girls (mean age 10.3, range 5.3 -
15.7 years), but this difference was not a statistically or clinically significant. The last three columns in Table 19 show the number of children from each diagnostic group, their proportion in the interview sample (n=24), and the proportion of each of the diagnostic groups in the larger study sample (n=200 plus three added for interviews only).

Table 19 Demographic Distribution of Interview Participants

<table>
<thead>
<tr>
<th>Diagnostic Group</th>
<th>Age in years Mean Range</th>
<th>Frequency Girls</th>
<th>Proportion of Interview Sample</th>
<th>Proportion in Study Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spina Bifida</td>
<td>9.2  6.4 - 12.6</td>
<td>6† (2)</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>Arthritis</td>
<td>10.3  5.3 - 12.1</td>
<td>6 (4)</td>
<td>25%</td>
<td>20%</td>
</tr>
<tr>
<td>Cerebral Palsy</td>
<td>9.8  7.4 - 13.0</td>
<td>4  (2)</td>
<td>16%</td>
<td>8%</td>
</tr>
<tr>
<td>Quadriplegia</td>
<td>15.8  15.7 - 15.9</td>
<td>2  (1)</td>
<td>8%</td>
<td>2% ‡</td>
</tr>
<tr>
<td>Muscular Dystrophy</td>
<td>15.2  14.8 - 15.8</td>
<td>3 (1)</td>
<td>13%</td>
<td>9%</td>
</tr>
<tr>
<td>Other</td>
<td>8.6  5.9 - 13.7</td>
<td>3‡ (2)</td>
<td>13%</td>
<td>36%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10.8  5.3 - 15.9</strong></td>
<td><strong>24 (12)</strong></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

† 1 boy initially referred as disabled, but global rating showed no limitations, thus was classified as non-disabled
‡ Arthrogryposis, Guillain-Barré, and a rare Neuromuscular Syndrome
§ 2% of study sample were amputees, and half of these had quadruplegia

Children with chronic conditions predominated the sample for two reasons. First, the children with acute conditions were enrolled in the responsiveness sub-component of the present research, therefore could not also be enrolled in the qualitative component (enrollment was limited to one additional sub-component to prevent an excessive burden on participating families). Second, during previous interviews with children, regarding physical disability, I found that those with chronic conditions had similar responses to those with acute conditions. However, the children with chronic conditions demonstrated greater insight and were more articulate. Thus, children with chronic conditions were thought to be most informative for this component of the research. Twenty-three of the children interviewed had chronic conditions. However, ten of the children with chronic conditions, plus the one child with the acute condition (Guillain Barre), experienced fluctuations in their disability over time.
At the time of their initial recruitment, the referring clinicians described seven of the children as having mild physical disability, 13 as moderate, three as severe, and one as non-disabled. Note that one non-disabled child (denoted by † in Table 19) was included to determine whether there were hidden disabilities or special issues that required attention to make the ASK appropriate to children with borderline disabilities. Thus, there were fewer with mild and no disability in this sample than in the larger study sample (n=200). The distribution of this sample according to severity of disability, as measured by the ASKp, ASKc and CHAQ, is shown in Table 20. The distribution of the larger sample (n=200) was previously shown in Table 7.

Table 20 Distribution of Disability Scores in Qualitative Sample

<table>
<thead>
<tr>
<th>Format</th>
<th>Scale</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentile†</td>
<td>ASK-p</td>
<td>65</td>
<td>24</td>
<td>26 to 99</td>
</tr>
<tr>
<td></td>
<td>ASK-c</td>
<td>73</td>
<td>22</td>
<td>34 to 100</td>
</tr>
<tr>
<td></td>
<td>Difference (ASKc-ASKp)</td>
<td>8</td>
<td>7</td>
<td>-2 to 27</td>
</tr>
<tr>
<td>Proportion‡</td>
<td>ASK-p</td>
<td>45</td>
<td>29</td>
<td>10 to 96</td>
</tr>
<tr>
<td></td>
<td>ASK-c</td>
<td>86</td>
<td>18</td>
<td>42 to 100</td>
</tr>
<tr>
<td></td>
<td>Difference (ASKc-ASKp)</td>
<td>41</td>
<td>22</td>
<td>4 to 73</td>
</tr>
<tr>
<td>Other</td>
<td>CHAQ</td>
<td>44</td>
<td>33</td>
<td>0 to 100</td>
</tr>
</tbody>
</table>

† Percentile scores are an aggregate based on 5 point response options.
‡ Proportion scores are an aggregate based on binary response options.

The children interviewed represented a broad range of disability levels and a broad range of discrepancies between capability and performance scores. Both capability-performance difference scores (percentile and proportion) were significantly different from 0 (p<0.0001), and similar to the larger population from which these children were selected.

1 The negative difference score for capability minus performance indicates one child with arthritis who regularly performed activities with difficulty. When the scores were re-computed based on binary coding (performed daily/not performed and capable/incapable) and reported as proportions, negative differences were not observed.
Exploration of the source of the discrepancy between performance and capability in the sample of children interviewed showed that the difference between ASKp and ASKc percentile scores was related to the ASKp scores (Pearson's $r=-0.42$, $p=0.058$) rather than with the ASKc scores ($r=-0.13$, $p=0.56$). The difference score based on proportion scores was also related to its ASKp component ($r=-0.79$, $p<0.0001$) rather than its ASKc component ($r=-0.08$, $p=0.75$ respectively). The correlation between percentile score differences and proportion score differences was significant ($r=0.68$, $p=0.0007$). These findings were consistent with those from the larger population of study subjects (Percentile: ASKp $r=-0.51$, $p<0.0001$, and ASKc $r=-0.13$, $p=0.069$) (Proportion: ASKp $r=-0.77$, $p<0.0001$, and ASKc $r=-0.07$, $p=0.353$).

Description of Family Composition

All the families lived within 120 km of Toronto, with two living in rural areas. Four of the families were single parent families, and four were single child families. The median number of children in the families was three (mean = 3.5, standard deviation = 4.4), with a range of one to 22. The largest family was the home of four children in the interview sample, therefore the remaining demographics are presented for the 21 study families.

Socioeconomic status (SES) and income status were computed from occupation titles using the Blishen et al. method based on the 1981 Canada census data [Blishen et al. 1987]. The mean SES based on the maximum SES within the household was 51.4 (standard deviation 10.2), and based on the average SES within the household was 47.9 (standard deviation 11.5). Both exceed the mean SES for Canadians in 1981 (mean 42.7, standard deviation 13.3) but only the maximum SES is significantly greater ($p=0.003$ and $p=0.080$ respectively). Mean income interpolated from the same index (maximum within the family) was comparable to the 1981 census data ($p=0.713$). In terms of the highest level of education among the parents; at least one parent had a university degree in four families, a college diploma in six families, some college or university experience in four families, complete grade 12 or 13 in three families,
some high school in four families. The lowest education level was grade 10, and the highest was a PhD.

Families were also classified according to whether or not they were associated with any disability organizations. Seven of the 21 families had no such affiliations. The children in these families were significantly less disabled (mean $ASK_p=79.8$, $ASK_c=85.8$) than the other children interviewed (mean $ASK_p=56.9$, $ASK_c=66.6$) ($p<0.04$). Four of these families had a child with arthritis, two had a child with diplegic cerebral palsy, and one had a child with S1 spina bifida. These parents stated they had not felt the need for such support. Those families that were affiliated with a disability group belonged to organizations such as the Easter Seal Society, Hot Wheels (wheelchair recreation program), the Muscular Dystrophy Society and arthritis support groups. These families tended to have a more severely involved child or be recently diagnosed and were therefore in greater need of support and/or information.

Overview

The 24 children interviewed gave many detailed examples of circumstances in which they had done specific activities, and circumstances in which they had not done the same activities or had experienced a problem with those activities. The specific activities discussed in each interview were primarily those that the child had reported some degree of capability to complete (on the $ASK_c$) yet had reported less than full performance (on the $ASK_p$). How an activity came to be either performed or not performed, when the child was capable of doing it, was thought to be the essence of the capability-performance distinction.

The children often expressed significant insight regarding how they made their decisions about the performance of activities. Older children were found to be slightly more articulate than younger children, and girls more articulate than boys. However, even the youngest boy contributed useful information to the qualitative research. Furthermore, the main findings of this qualitative research were consistent across age, diagnostic groups and gender as is demonstrated by the inclusion of all groups and both genders in the quotes presented.
Analysis of children's accounts of the differences between capability and performance revealed three major themes that were common among the quotes from the 24 children interviewed. These were: *affinity for an activity*, *how formidable the activity was perceived to be*, and *modifications*. Children's performances were the result of an integration process that considered these themes. The themes were often considered concurrently by the children. However, the themes were sometimes considered sequentially, in that status with respect to one theme (e.g., negative *affinity*) precluded further consideration of the activity or required consideration of the next theme (e.g., if the activity was *perceived to be formidable* then consideration of *modifications* was required). The themes are presented here in an order that reflects this sequence.

The theme *affinity for an activity* relates to children's accounts of deciding whether or not to bother doing an activity because of likes and dislikes (e.g., do not like running games). If the child disliked an activity it was sometimes excluded, regardless of the other factors. Thus, *affinity* was considered first. However, if the child liked the activity, how *formidable* the activity was perceived to be, was considered next. Formidability was defined as how challenging the activity was to perform. If a task was considered only a little challenging, there was often no need to consider modifications. The child simply completed the activity by the usual non-disabled method albeit with a little problem (where problem included extra effort, time or adverse consequences such as discomfort). When the activity was both valued by the child (*affinity*) and was considered *formidable* then the third theme, *modifications*, was also considered. The *modifications* theme was subdivided into: use of assistive devices, alteration of technique, and active use of help.

In addition to the three themes identified, there were several elements (defined as factors of lesser impact than the themes) within each theme that also contributed to performance. These elements (identified in children's accounts) included: characteristics of the internal environment, characteristics of the external environment, quality and efficiency of the performance, desire for autonomy, and the consequences of performance. Many of these five
elements pertained to all three themes. However, the children’s accounts did not necessarily implicate all elements for all themes.

The elements were broadly defined and were related to one another. Characteristics of the internal environment included the children’s considerations of aspects inherent to themselves that influenced their performance of activities (e.g., confidence, current symptom level, motivation). In this analysis it was the distinction between capability and performance that the analysis was concerned with (i.e., interested in factors beyond physical ability that may interfere with performance). Characteristics of the external environment included consideration of all aspects of the physical and social environment relevant to the child’s situation (e.g., terrain, season, presence of peers). Quality and efficiency of the performance included concerns with the expected excellence of the activity if performed (e.g., normality of gait pattern and overt evidence of exertion). Desire for autonomy included the strength of their desire to be independent with any given activity (and was found to increase with increasing age). The consequences of performance covered the children’s considerations regarding the costs they anticipated having to pay as a result of performing an activity (e.g., in terms of fatigue or physical discomfort).

A review of the three themes and various elements across various activities and diagnostic groups showed that these children went through a trading-off process in determining their performance of activities. The three key themes and their associated elements interacted in this trading-off process and exerted both positive and negative influences. For example, the availability of an assistive device may deter a child from accepting help (e.g., braces improve balance and negate the need to hold Mom’s hand), or conversely may necessitate the active use of help (e.g., braces require Mom’s assistance to put on). A discussion of each of these themes and related elements follows. Specific quotes from the children are used to illustrate how the competing themes and elements were integrated in the children’s decision making process regarding the performance of activities.
The quotes selected are representative not only of the accounts of that child, but speak for all the children interviewed. To support this assertion, quotes from a variety of different children have been used. Each quote is marked with the age, gender and diagnostic group of the child to illustrate the consistency of themes and elements across the children who were interviewed. To provide a sense of consistency for the use of the themes and elements across activities, a synopsis of the accounts of two children across a variety of activities has been included at the end of the qualitative results. These two vignettes support the consistency of themes and elements across activities. The pertinent literature is integrated to support the generalizability of results beyond this study population. This integration was reserved to the end of the chapter to prevent the literature from influencing the interpretation of children’s accounts, in accordance with the principles of grounded theory [Strauss and Corbin 1990].

Figure 8 Overview of Trading-Off Process
Themes

The themes and elements shown in Figure 8 are presented with supporting quotes in the following sections. The concept of trading-off and the three main themes were consistent across children and across activities, however, the same is not true of the elements. The presentation of children's accounts documents where particular elements were observed primarily in particular sub-groups.

Theme I  Affinity for an Activity

Review of the interview transcripts revealed that children with disabilities are similar to their non-disabled peers in their affinity for certain activities; they have activities that they like and dislike. Just as non-disabled children may not like tag and therefore not play tag, children with disability may not like and not play tag. And similar to their peers, these likes and dislikes may be influenced by their capability. Children will opt out of an activity they are terrible at, if given the option. Thus, affinity for certain activities may explain why some activities are performed and others are not, even before elements (e.g., the situation) are evaluated.

The distinction between affinity and disability regarding performance decisions was far from straightforward. This was, in part, because children focused mainly on optional activities such as play and getting snacks when they talked about affinity for activities. Sometimes the children attributed lack of performance of an activity to 'dislike,' when it may have alternatively (and perhaps more appropriately) been classified as disability. For example, one 10 year old girl with mild disability due to arthritis discussed her dislike for sports. She talked about her weekly 30 minute gym session with great disdain, and referred to it as her "least favorite" class. The following insightful statements suggest that she was aware of the difficulty distinguishing between her dislike of the activity and her disability. In the account the girl articulates the possibility that, if her disability did not limit her performance of sports, she may or may not like sports. She suggests that because she has not had much practice at
running (due to disability), that she is not good at it, and that this has led her to dislike running. Therefore, her dislike of sports may indirectly be due to her disability.

Quote 2-1-3 (10 year old girl with arthritis)
Child: Maybe it's just because it doesn't interest me, or I'm not very good at it because I don't run fast or whatever, but I just don't do a lot of it so I don't really know if I'm good at that or not, depending on my arthritis or whatever, I don't know. [pause] If, like if I did it, I don't know if that would really interest me that much.

This girl also denied that she had arthritis, and frequently attributed her limitations to preference rather than to arthritis. Later in the discussion she returned to this issue of distinguishing between affinity and disability.

Quote 2-1-3 (10 year old girl with arthritis)
Child: How can I say; I was never really energetic but I'm not now so I don't know. Maybe that's just because I can't now and maybe that's just because I don't want to now. If I can't I don't know why really, because, like there are a lot of things.

Still later in the interview, this girl revealed that her dislike of sports is unusual in her family, just as her arthritis is unusual in her family. This further suggests that her dislike was a form of "disguised" disability.

A 12 year old boy, classified as having no disability in the presence of mild impairment (S1 spina bifida), confirmed the connection between dislike and disability made by the 10 year old girl with arthritis. He described less difficulty roller-blading than running, that was associated with a greater liking for roller-blading than for running.

Quote 11-1-4 (12 year old boy with S1 spina bifida)
Child: I don't run as fast. I run out of energy. Like I can roller-blade for like hours but I can't run.
NY: What's the difference?
Child: I don't know. [pause] I don't know, I don't like running that much. I can roller-blade as fast as I want.

The power to roller blade comes from muscles unaffected by S1 spina bifida, whereas running requires repetitive use of muscles weakened by S1 spina bifida. Thus, this boy's minor running disability may be one source of his lack of affinity for running.

Based on these data, the issue of affinity for activities appears to be complicated. Affinity may be a reflection of true dislike (an innate expression of values independent of abilities) or a consequence of disability. Regardless of the source, lack of affinity for an activity may create a capability-performance gap. For this reason, the response options on the ASKp questionnaire
were left open to the children’s interpretations for activities in which dislike may be common (e.g., the leisure activities question gave examples, but was not restricted to one activity such as bicycling). This allowed children to report their performance of their usual activities (ones they liked), and avoided coding children as disabled because of dislike alone. On other questions not-applicable response options were included to prevent true dislike from contributing to their disability scores. Because of these provisions, the capability-performance gap observed on the ASK may be attributed to disability-driven affinity, primarily.

Elements Related to Affinity for an Activity Theme

As mentioned in the overview, all the themes were influenced by a series of sub-themes or elements. These elements represent an additional level of depth in the analysis, and were not observed with as much consistency across children and activities as were the three main themes. However, the elements remain important in understanding the trading-off process.

The children talked about two elements with respect to affinity; those related to the internal and external environment. For example, children with arthritis talked about how affinity for an activity was altered by symptoms. Symptoms are one aspect of a child’s internal environment. One 11 year old boy described how his arthritis fluctuated from day to day, and how he did not like to play some games when his arthritis was bothering him.

Quote 3-1-11 (11 year old boy with arthritis)
Child: Um . . . today was a good day ‘cause it didn’t bother me too much. But sometimes, um, like it bothers me a little bit.
NY: Do you still play tag on the days that it bothers you a little bit?
Child: No.

The same child also talked about how elements of the external environmental, specifically the presence of peers altered his affinity consequently and his performance. Specifically, that some of the likes, dislikes and choices regarding activities were not his alone but belonged to his peer group.

Quote 3-1-11 (11 year old boy with arthritis)
NY: Did you play tag today?
Child: No.
NY: No?
Child: ‘Cause um, ‘cause we didn’t feel like it. We were just talking.

...
NY: How do you decide if today is a tag day or not?
Child: Like it depends on the weather. If it’s cold, then we don’t play.
NY: And is that your choice or is that everybody’s choice?
Child: Everybody’s choice. (laughs) ‘Cause they say if it’s too cold, and you run and you want to take off your jacket and you take it off, then you get a cold.

This example shows that at least two aspects of the external environment (cold weather and peers’ affinities) influenced his performance.

An older boy with muscular dystrophy also suggested that his dislike of cold weather (an external environmental factor) and his affinity for play activities may be connected. He usually liked to go out to play regularly, however, in the cold weather he played infrequently because of the cold.

Quote 20-2-1 (15 year old boy with muscular dystrophy)
NY: Playing outside with your friends; you said you do sometimes.
Child: Um hm. Mostly they come to my house and play games. I usually don’t go outside, especially now that it’s cold. But in the summer I go out.

... Child: I plain just don’t like it [the cold, and therefore prefers not to go out and play].

As a result of his dislike for cold, this boy did not like to play outside in the winter, and experienced greater disability in the winter time. This dislike and associated disability may be directly related to his muscular dystrophy, because cold is often experienced more intensely by children with muscular dystrophy.

These examples illustrate that when dislike for an activity was strong and negative, it alone dictated the decision not to do an activity. Thus, affinity was the primary consideration in the decision of whether or not to perform an activity.

**Degree of Affinity**

So far the examples have described affinity as having a yes/no effect on the choices that children made regarding their performance of activities. However, when there was a state of neutrality or positive affinity for an activity the possibility of completing the activity was entertained by the children. Under these circumstances the degree of affinity (rather than just the presence or absence of affinity) became an important factor in the trading-off process, along with consideration of how formidable the activity was and the availability of modifications.
Children expressed varying degrees of affinity (or preferences) across a range of activities, and were more motivated to find solutions to achieve activities that they liked. For example, one girl talked about one of her favorite school trips, for which she had to arrange her own attendant and give up the independence of her power wheelchair because it could not be transported on the school bus.

Quote 16-2-9 (15 year old girl with congenital muscular dystrophy)
NY: Does your mom go with you on school trips or your dad?
Child: No.
NY: That’s a strategy some of the younger kids have mentioned, that mom gets to be the helper on the trip so they don’t have to go on their own. You’d rather not go?
Child: Yeah.
NY: Do you opt out of the school trips a lot and just say that no you’re not going to go?
Child: It depends where they are going, and if they can take my wheelchair on the bus so I can get around.
NY: If they go somewhere and you really like it, do you try to go no matter what barriers are put in front of you?
Child: Yup.
NY: What kinds of places do you like to go to? What was the best trip last year?
Child: We went to Quebec.
NY: You went to Quebec? Wow, that’s quite a school trip.
Child: Yeah, we went for four days but they couldn’t arrange for anybody to accompany me.
NY: So your sister went to help? I bet she liked that.
Child: Not really. She kinda got muscle pain because she had to carry me on and off the bus so many times, and I took the manual [wheelchair] so she had to push me.
NY: Would you do it again?
Child: Yeah. I really like Quebec.
NY: Would your sister do it again?
Child: Probably not. (Both laugh.) Not with me at least.

Her desire to go to Quebec appears to have been a strong motivator for this individual. She elected to trade away independent control over mobility in order to go on the trip.

The example above also shows that degree of affinity was not only associated with specific activities and places, but also with methods of completing activities. Thus, children expressed preferences related to assistive devices, alteration in technique, active use of help, and use of alternate transportation modes. For example, one of the girls with diplegic cerebral palsy (Quote 13-1-3) stated that to get around “I walk and sometimes I crawl.” In fact, she described crawling only at home and preferred to walk or use a wheelchair at school. Thus, this child preferred to be upright at school and would trade-off the expenditure of extra effort (a
consequence of performance) in order to walk, or accept the limitations of the wheelchair rather than crawl.

Unlike the examples in which affinity was very strongly negative and was the sole reason not to do an activity, the role of degree of affinity as one of many factors was rarely stated explicitly. Often its role could only be inferred from reported behaviors in various environments. However, through many examples it became apparent that degree of affinity issues had a very broad scope, so that strong preferences in one area sometimes affected disability in another area. For example, a nine year old boy with T12 spina bifida stated that he was very particular about the cleanliness of his manual wheelchair, and simply preferred not to get it dirty (Quote 6-2-1). This child would forgo the opportunity to play outside for the preferred state of a clean wheelchair. Two other children in power wheelchairs had similar accounts, and explained that their outside mobility was restricted in bad weather because they preferred not to drag mud into the house too often and inconvenience others.

Quote 16-1-17 (15 year old girl with congenital muscular dystrophy)
Child: I try not to go in the snow because it is all wet, and the hardwood shouldn’t get wet or it will ruin. And I can dry it [the wheelchair] off but that takes a while, so I kind of stay out of the snow. My Dad shovels it and tries to get it [the driveway] all clean [so that she doesn’t drag in too much snow].
NY: So that’s why you prefer to stay inside?
Child: Yup.

The importance of cleanliness and the feelings of others were echoed by another child who used a power wheelchair.

Quote 22-2-6 (15 year old boy with quadriplegia)
Child: Well I don’t want to bring a mess because usually the floor is clean and somebody has to mop it again.

This latter quote came from a boy who lives in a very large family, and lacks the mobility to clean his chair or the floor himself. His performance of outdoor activities was driven by his heightened awareness of the distribution of chores within the family, and his reluctance to make extra work for others.

Another child discussed how his decision to enroll in a school that had a strict uniform code made it impossible for him to manage getting dressed for school independently.
Discussions with the child in the above quotation and with his mother suggested that the boy’s preference for this particular school was a direct reflection of his mother's strongly held preference. Thus, the preference was not entirely the child’s. Another boy talked about his experiences at a respite care centre, where the staff prefer to do much of the morning care for the children.

Finally, there were times when the preferences of others superseded those of the child. Because of the preferences of others, the children do not have the opportunity to engage in some daily care activities, their performance scores for these activities drop, and the gap between performance and capability is increased.

Both of these examples show how others’ preferences figure into the trading-off process. In summary, the role of affinity for activities in the trading-off process can be seen in the many examples in this section. If children do not like an activity, it is dismissed (with the exception of mandated activities). If they do like it, they go on to assess the magnitude of their preference or liking of the activity, and often they also consider the preferences of others to determine whether or not they will do an activity.
It can also be seen in the previous examples that these factors have little effect on ASKc scores, because regardless of affinity or preference they either can or cannot do an activity. However, these factors do have a significant effect on disability as measured by the ASKp. For example, the 15 year old boy with muscular dystrophy would score lower on the ASKp during weeks in respite care, because of the preferences to the centre’s staff. The boy who chose to attend a school with a uniform requirement scored more disabled because of that preference. This degree of heightened disability is experienced by the child, and detected by the ASKp but not by the ASKc. Thus, affinity for activities contributes to the discrepancy between capability and performance. Therefore, this factor explains why there is a greater capability-performance gap evident in activities children dislike or have low affinity for (opt out readily) versus those that children have a high affinity for. If they really like an activity and have the capability to do the activity, they will find a way to perform the activity.

The affinity for an activity is specific to that activity, and when considered in isolation (i.e., free from the influences of peers and other motivators) affinity for an activity is relatively stable over the period of a week (the ASK’s reference time frame is one week). But what about factors that are not specific to the activity, or not stable over a one week period? How formidable the activity is perceived to be is one such factor.

Theme II  How Formidable the Activity Was Perceived To Be

An activity was considered formidable or challenging when there was a mismatch between the perceived demands of the activity and the perceived abilities of the person. When a child considered an activity to be very formidable, they frequently did not perform it, or at least not in the typical non-disabled way. Conversely, when a child perceived an activity as not formidable at all, they frequently performed it by the usual non-disabled method. Based on the interview data, the magnitude of the perceived challenge was dependent upon the demands of the activity and the ability level of the person.
**Activity Demands**

Activities vary in their degree of difficulty. The varying difficulty levels for the activities included in the ASK were identified by Rasch modeling, and presented earlier in this chapter. For example, climbing stairs was more difficult than walking indoors. The more difficult an activity was the more likely it was to exceed the ability level of the child who needed to complete the activity. Children used the word ‘hard’ to express the high difficulty level of an activity. Their meaning of ‘hard’ was physically difficult. This can be seen in their accounts of how cursive writing was ‘hard’ in comparison to printing and in their accounts of physically demanding activities such as bed-making. A typical example is shown below.

Quote 23-1-3 (10 year old boy with hemiplegic cerebral palsy)
NY: What kinds of things do you need help with?
Child: I need help [long pause] well help making my bed.
NY: Can you do it by yourself?
Child: Yeah but it’s just hard.
NY: It’s just hard?
Child: Yeah!
NY: Are you tired when you’re done?
Child: I’m frustrated.

This example shows the meaning of hard, and also how the demands of the activity caused him to trade-off in favor of accepting help.

**Ability Level**

Similar to the difficulty level of the activity, the Rasch model also identified the ability levels of each child. As was shown in Table 20, the children who participated in these interviews had a wide range of ability levels. Children with extensive limitations in ability (e.g., due to T12 spina bifida or quadriplegia) were frequently unable to complete activities, while children whose ability levels approached those of their non-disabled peers completed activities more often.

As was seen in the example of the 10 year old boy’s bed-making behaviour, the ability level of the individual interacted to a great degree with the difficulty level of the activity. Because this activity was close enough to his ability level that he had the capability to do it, but
not so close that he could do it easily on a daily basis, this activity increased the capability-performance gap for this child.

Another boy described a similar trade-off with respect to bicycling. Although he had almost completely recovered from his acute neurological condition, his legs remained slightly weak, and his balance reactions were a little delayed. As a result, his ability level was very close to the demands of bicycling. His mother stated that he had not been on his bicycle since his illness.

Quote 17-2-9 (5 year old boy with Guillain-Barre)
Mom: Why don’t you want to ride it when we took the training wheels off [as per the child’s request]?
Child: Scared.
NY: What are you scared of?
Child: Going boom.
NY: Going boom? Have you gone boom a lot?
Child: [Shakes head no]
NY: No? A lot of kids fall down a lot when they’re learning again.
Mom: Well, he won’t put himself in situations where he might get hurt.
NY: Now is that different for him from before, or has he always been a safety conscious child?
Mom: Not as safe as that.

The closeness of the activity’s demands to the child’s ability level resulted in uncertainty regarding performance. When he considered the possible consequences of trying to ride his bicycle (falling), the additional element swayed the trade-off away from bicycling.

The same interaction had a more favorable outcome for a seven year old boy with diplegic cerebral palsy (Quote 14-1-5) whose ability to play computer games far exceeded the demands of the game. This boy did not require any modifications nor did he suffered adverse consequences related to this play activity, and therefore he showed no discrepancy between capability and performance.

In summary, children were usually unsuccessful in completing activities well above their ability levels (i.e., relatively hard activities such as in the bed-making example), demonstrated variability in performance of activities very close to their ability level (i.e., closely targeted activities such as in the bicycle example), and were usually successful in completing activities well below their ability levels (i.e., relatively easy activities such as the computer game
example). It was in situations of uncertainty that the influence of other themes and elements had the greatest effect.

Elements Related to Formidability Theme

As was discussed for affinity, how formidable an activity was perceived to be was determined, in part, by many elements. Primarily, these included aspects of the situation. The term ‘situation’ is used to encompass both elements related to the individual and to the external environment. Some of the key elements related to the individual included: state of symptoms, and motivation. Some of the key elements related to the external situation included: the physical environment (e.g., type of floor surface), time pressure, and presence of help or assistive devices.

Because the definition of formidable or challenging includes the perceptions of the individual, internal environment elements that influenced the individual’s perceptions were considered. The most frequently discussed of these elements was the impact of current symptom level. The children talked about how their symptoms of fatigue, hurting, restricted range of motion, shortness of breath, stiffness and weakness affected their perceptions of how challenging activities were, and how these elements determined their performance. For example, one girl who had a choice between multiple assistive devices related her choices to fatigue.

Quote 13-1-5 (13 year old girl with diplegic cerebral palsy)
NY: So how do you decide; wheelchair or walker?
Child: If I’m tired I use my chair, if I’m not I use my walker.
NY: So does it depend on what time of day it is?
Child: No. Just whenever I get tired.

Because of the nature of their disabilities (e.g., arthritis) some children’s symptom levels were associated with the time of the day. Thus, time of day altered the decisions children made about activities and modifications of those activities.

Quote 1-2-7 (10 year old girl with arthritis)
NY: If I said let’s go for a little run right now, would you go? [Very long pause] I brought my running shoes?
Child: [Very long pause] um [pause] um [pause]
NY: How do you decide? What do you think about?
After symptoms, the next most frequent internal element was motivation. Much of the influence of motivation has already been addressed as an element under the heading of affinity, and is mentioned here only to add emphasis. Another child also gave an example of running. Similar to the previous girl’s example, he found that it hurt. However, unlike in the girl’s account, this boy did run when he was motivated (e.g., during tag with his friends).

Quote 3-1-8 (11 year old boy with arthritis)
NY: You say sometimes when you run fast, it hurts. When did you run fast last?
Child: Um [pause] we were playing tag and running, and I started running fast [oblivious of pain]. And when I touched someone, then [after the goal had been fulfilled and the distraction ceased] it started hurting then they, and then it started hurting, then they, and then um [pause], then I ran again, ran slowly and it didn’t hurt that much.
NY: So if you go slower, it’s a little better?
Child: Yup.

In this example his motivation and enjoyment of playing with his friends blocked out the negative influence of pain, until he achieved his objective. Then his motivation dropped and his awareness of pain returned. Only then did he modify the activity to reduce his symptoms.

Lack of motivation also had an effect. Two children talked about lack of motivation and “laziness” as an element related to their non-performance of transfers.

Quote 13-2-10 (13 year old girl with diplegic cerebral palsy)
NY: Getting in and out of bed; you were doing by yourself sometimes [according to ASKp booklet].
Child: And sometimes my mom just pulls me up ‘cause I’m lazy.
NY: Does she pull your sister out too?
Child: No, ‘cause ‘cause she’s on the top bunk.

Quote 16-2-7 (15 year old boy with congenital muscular dystrophy)
NY: Getting in and out of the chair; you do on your own most of the time. [Pause] What’s special about the few occasions that you don’t do it on your own? [Very long pause] Mom does it, why does mom do it?
Child: Sometimes I’m lazy.
NY: Are you lazy or wiped out?
Child: I’m just lazy.
NY: Ok, that’s honest. That’s good.
These examples show that these children are apparently no different from most people, who forgo activities from time to time because of "laziness".

Besides the pertinent aspects of the child's internal environment (i.e., activity demands of the item, the ability level of the child, symptom levels and motivating elements), there were external environmental elements that also contributed to the trading-off process and the difference between capability and performance. These included aspects of the physical environment and time pressure (or time constraints).

With respect to aspects of the physical environment, the children talked about environmental barriers. Frequently these were accessibility issues, such as objects in the way at school and doorways too small to permit walkers to pass through.

Quote 12-1-3 (8 year old girl with diplegic cerebral palsy)
NY: Ok. So tell me about the walker at school.
Child: Umm. [very long pause] I like to walk in it but it's hard to get around when the walker is right beside a big box and I have to get in between the walker and the box at the same time just to get my bag.

Child: Whenever I can't fit through the washroom door with my walker so they have to walk me to the washroom without my walker.
NY: So there are a couple of places your walker doesn't go at school.
Child: Yeah.

Other examples related to limitations in accessibility at friends' houses and in their own home.

Quote 9-1-4 (9 year old boy with T12 Spina Bifida)
NY: One of the things you told me [in the ASKp booklet] you had trouble with was turning on the taps in the morning.
Child: Yeah.
NY: What happens when you try to do the taps in the morning?
NY: It's tricky, I bet. What's tricky?
Child: That it's high, and there's no room at the sides [of the sink].
NY: So there is not enough space for you?
Child: Yeah.
NY: Is it hard to maneuver in the bathroom?
Child: Yeah.
NY: Does somebody help you in the bathroom?
Child: Mom or Dad.

Although this boy could manage the taps, because they are situated in a small bathroom he always had a parent to help him. Thus, the size of the bathroom contributed to his performance
disability, and the difference between capability and performance. In this situation there was little to consider. Thus, he traded away his independence to avoid a long and cramped struggle in the bathroom. In a related example, a 15 year old girl with congenital muscular dystrophy described the barriers of front porch steps. She found this to be a formidable barrier to access for her wheelchair, since all the neighborhood homes had at least one step. She did not experience this limitation at home because of substantial modifications. As a result she rarely went to her friends’ houses to play.

Although there were many examples of physical barriers in the environment, there were fewer barriers than expected. This was due primarily to environmental adaptations that had been made to many of the homes (e.g., installation of elevators) and schools (e.g., automatic doors) to accommodate the children’s needs. These adaptations will be addressed in more detail in the section on modifications. There were, however, many other less obvious environmental elements that were identified by the children. These included seasonal problems (i.e., ice and mud), terrain problems (i.e., gravel and carpet), and the effect of crowds (i.e., reduced confidence and challenged balance). In the quote below, some of the environmental factors are listed.

Quote 5-2-3 (11 year old girl with arthritis)
NY: Is it harder to do things on the gym floor than it is outside?
Child: Actually it’s easier.
NY: Easier. Why is it easier?
Child: Because outside it’s bumpy and there are rocks. Half of our soccer field is rocks and grass so it’s like pavement and then rocks all over it.
NY: So it’s like gravel?
Child: It’s gravel actually. (Both laugh) And um, so when we’re running on it and you stop, your foot can slide and you can fall flat on the floor on the rocks so it’s easier in the gym because you don’t slide . . . you stop.

Thus, gravel increases the difficulty level. Similarly, crowds increase the difficulty level for some children.

Quote 12-2-4 (8 year old girl with diplegic cerebral palsy)
NY: Do you have any trouble with crowds?
Child: No. Only at school. The people in line keep pushing me over and I [yells and simulates a fall].
NY: Do you think you fall over easily?
Child: Yeah. I even fall over if someone accidentally kicks me in the foot.
Another aspect of the external environment that was frequently discussed was \textit{time pressure} or time constraints. This is a separate issue from time of day that was discussed previously as an internal environmental element related to diurnal symptom patterns. The children discussed time pressure as a constraint within which they were expected to complete an activity. For example, on school day mornings children are expected to dress, eat and get to school in a short period of time. In this situation, time is an external environmental element. In a related example one girl talks about the extra time required for her to get in an out of the car independently.

\begin{quote}
13-2-3 (13 year old girl with diplegic cerebral palsy)
NY: What about getting in and out of the car by yourself? You say you don't do that.
Child: No, my parents lift me in.
Brother: Or your big brother.
NY: Or your big brother?
Child: Well now, we switched, and actually, sometimes I get in now. Now I can get in.
NY: What's changed?
Child: I don't know. I just kinda [long pause] my mom wanted me to start trying to do that. She still lifts me in and out sometimes. It kind of matters [long pause].
NY: Does it take you a long time to get into the car?
Child: Sometimes.
NY: So do you think maybe that's why mom lifts you in?
Child: Yeah.
NY: Cause it's faster?
Child: Uh-hmm.
\end{quote}

This girl, like many of the children, was aware that taking a long time to complete an activity was a burden on other peoples schedules and that the act of lifting her into the car was almost equally burdensome. As a result her behaviour shows an oscillation between trying not to take too much time (accepts assistance) and trying not to overburden her caregivers (transfers independently).

The importance of time limitations was discussed most explicitly by children with muscular dystrophy. For example, a very independent girl with advanced muscular dystrophy discussed the time element related to dressing.

\begin{quote}
16-1-7 (15 year old girl with congenital muscular dystrophy)
NY: And what happens on the days that you do get your socks on? Can you ever get them both on?
\end{quote}
Because there was only one aspect of dressing that was particularly problematic, she traded for partial assistance with her socks and continued to put on the remainder of her clothes independently.

A boy whose muscular dystrophy was much less advanced had similar comments regarding time. His time constraints were much tighter than the previous girls, and his comments went further to describe how time constraints prevented him from even partially completing many activities independently.

Quote 19-1-4 (15 year old boy with congenital muscular dystrophy)
NY:  Let's talk about some of the other things that you do some of the time and figure out what some of the issues are and how you choose. One is getting on and off the toilet. You do it most of the time. What happens the rest of the time?
Child:  The rest of the time again we're sort of in a rush and you just can't struggle and you have to go.
NY:  So if you had an endless amount of time, you would do it all the time?
Child:  Yeah. The thing is, for most of these answers, it is technically time that's one of the bigger things. It's someone else's schedule you're working on.

For this boy, the predominant limiting element in his life was time restrictions. He could have accomplished most activities if time had permitted. What was particularly noticeable in the interview with this boy was how he attributed the time limitations to others, as was alluded to in the quote above, and is explained in the quote below.

Quote 19-1-4 (15 year old boy with congenital muscular dystrophy)
Child:  Yeah with transportation you really are at Wheel-Trans' mercy, you have to sort of follow things, you just can't like if they are there early you just have to go. Because there are other passengers as well and that sort of problem, because you have to go, because those buses are big, and they're blocking traffic. And also for the taxi drivers, it's hard for them because they have a schedule, and like some people think you have someone here [to provide physical assistance] when you really don't.

Thus, both the physical and temporal aspects of the external environment affect performance, and may contribute to the capability-performance disparity.

One final aspect of the physical environment in need of mention is the presence of help or assistive devices. While modifications are a key theme, and are discussed in the next section (Modifications/Resources), their presence or absence may also be considered part of the physical environment. The presence or absence of assistive devices was discussed less
frequently than were physical barriers and time pressures, however, children mentioned that
sometimes they did not do an activity themselves because there was an easier alternative. For
example, Mom was already there to put her braces on, so why not let her do the socks too?

Quote 13-2-2 (13 year old girl with diplegic cerebral palsy)
NY: [Reads from booklet] Once in a while you do your socks.
Child: Once in a while, 'cause usually my mom will do my braces and that's when she does my socks and
then she puts my braces on.

In another example a girl reported that she used an alternative mode of transportation
because it was readily available.

Quote 21-2-10 (15 year old girl with quadriplegia)
Child: At school I use the elevator but I could use the stairs. I'd be late for class. There's an elevator
there so why not use it.

In summary, how formidable an activity was perceived to be was determined by the
demand of the activity, the ability level of the person, and the characteristics of the situation.
Analyses of the interviews revealed that when the activity was formidable, and an assistive
device or person was available in the environment, the use of the resource was more readily
considered than had it not been available in that environment. How these resources participated
in the trading-off process is considered below.

Theme III Modifications/Resources

Part of the variability in performance, and how formidable an activity was perceived to be,
was related to the availability of alternative options and resources to modify the activity.
Children showed through example that they modify activities via the: use of assistive devices,
alteration of technique, and active use of help. Assistive device use included all types of
devices (e.g., ankle-foot-orthoses, wheelchairs and elevators). Alteration of technique
describes variations on how the activity was done (e.g., go more slowly). Active use of help
describes the elements that relate to recruitment of someone else to work with them to complete
an activity. These three components were not considered in any consistent order by the
children in their trading-off process, but these alternatives played a major role in modifying the
degree of difficulty of activities.
Use of modifications and resources varied across different activities, but also varied within the same activity. For example: a cane may have been used in walking, but not in playing and stair climbing activities, and its usage may have been variable during walking. Thus, while factors such as affinity belong to the individual and were consistent over a period of time, modifications and resources may be extremely variable over time and their impact much more complex.

**Assistive Devices**

When the children described an activity in which they had variable performance, one of the elements that differed between the situations in which they did perform versus those in which they did not perform was assistive devices. Among the types of assistive devices they talked about, were personal assistive devices, environmental modifications, and alternative modes of transportation. *Personal assistive devices* are worn by the child and facilitate activity and include ankle-foot-orthoses, resting splints, protheses, walkers, pencil grips, standers, orthopaedic shoes, spinal braces, and buckets. *Environmental modifications* are attached to a physical environment, rather than a person, and include custom bathroom fixtures and wheelchair ramps. *Alternative modes of transportation* are distinct from personal assistive devices because they altered the activity by replacing it with a less demanding alternative (e.g., disabled transit to get to school) or with a completely different activity (e.g., a manual wheelchair that changes walking into an upper extremity function), rather than altering the child's approach to the activity. These alternatives included wheelchairs, scooters, kiddie cars, and bus systems for the disabled.

As mentioned earlier in this section, the use of modifications was variable even within the same child for a very similar activity. For example, one girl described the variable use of an assistive device for writing.

*Quote 5-3-6 (11 year old girl with arthritis)*

NY: Do you have any tricks to make it easier when you are writing?
Child: No. I can't skip anything.
NY: Do you use any special pens or anything?
Child: Umm. I use a pen with a gripper. I have one of those triangle things that are for your pencil. I only use that for math because I have a lot of math.
Assistive devices were more often used because they enhanced function (decreased the formidable nature of the activity), because they reduced the likelihood of adverse consequences (e.g., improved safety), or because the child had been told to use the assistive device by a clinician.

During the interviews it became apparent that the use of assistive devices altered disability and thereby altered the capability-performance discrepancy, but not always in a positive direction. Assistive devices were found to either increase, have no impact, or decrease the gap between capability and performance. Most commonly the assistive devices decreased the gap, by enhancing performance. For example, children reported that using a wheelchair increased their mobility and made it possible for them to keep up with their peers.

Quote 8-2-4 (7 year old girl with T12 spina bifida)
NY: Can you keep up with your friends at school?
Child: Yeah. I can go faster with my chair.

This girl noted a great improvement in efficiency, that made the wheelchair a favored assistive device for her.

However, the children also spoke about situations that were not amenable to the use of assistive devices. In these situations assistive devices were ineffective in diminishing the gap.

Quote 12-1-3 (8 year old girl with diplegic cerebral palsy)
NY: What do you do with the walker when you get to the stairs at school? You said you wear a helmet on the stairs.
Child: Uh, I fold it up and I go up the stairs. I walk down the hall to the library and I sit down in a big chair. Then at the end of library, I go back to my classroom and get the walker.

Children also gave examples of assistive devices that removed one barrier, but created another challenge in the process. In these situations assistive devices were partially effective in diminishing the gap. For example, the boy in the following quote has an elevator to enable him to get to his bedroom independently, but because of the design of the elevator, it takes significant effort on his part to use it.

Quote 9-2-7 (9 year old boy with T12 spina bifida)
NY: Tell me about your elevator. How long have you had your elevator for?
Child: Two years old.
NY: Is it good?
Child: Way better than last time.
NY: What's the difference?
Child: Mom and dad had to lift me upstairs when we didn’t have it.
NY: So they had to carry you upstairs?
Child: Yeah.
NY: Did you like that?
Child: Uh-huh.
NY: So how come you like the elevator then if you liked being carried so much?
Child: Why did you get me the elevator?
Mom: Well, because you wanted to do it yourself didn’t you? You wanted privacy so you could take off.
NY: Would you rather take the elevator to go to bed tonight or would you rather someone carried you upstairs?
Child: Carried.
NY: Really?
Child: I’m just tired tonight.
NY: Oh. Does it take a lot of work to get in the elevator?
Child: Yeah a little.
NY: Can you do it all by yourself?
Child: Yeah.

It was apparent upon observing the child, that the two centimeter ramp at the entrance required special skill. This was not because of the size of the ramp, but because the child had to exert enough force to get up the ramp then quickly stop, because of the small space inside the elevator.

In other situations the assistive devices that were selected for one activity became a barrier and reduced performance of another unrelated activity, thereby increasing the gap. For example, splints were important for reducing the capability-performance gap in walking, but caused the need for assistance in dressing.

Quote 13-1-13 (13 year old girl with diplegic cerebral palsy)
NY: You don’t always do your pants by yourself, you do it sometimes [reported in booklet].
Child: Because of my braces. I have to put my braces on before my pants.
NY: So having braces on makes it a little harder to get your pants on?
Child: Uh-huh. That’s why I want to quit wearing braces.
NY: Oh.
Child: So then I can put my pants on easier.
NY: What happens if you walk without your braces?
Child: Well, I can walk but my ankle will start hurting because it doesn’t have enough support.

It was in this latter type of situation where trading-off was most visible, since the child had to consider whether the benefits provided by the braces (e.g., better support and less pain) were worth the disability they caused (e.g., in dressing).

The trading-off process and the impact of assistive devices on the capability-performance discrepancy were also very visible in the choices children made between multiple mobility
devices. Of the 24 children interviewed, seven used no mobility devices regularly (including five children with arthritis) and three were obligatory wheelchair users. Thus, the remaining 14 children had one or more options available to them. These 14 children represented primarily the mid range of the disability continuum. Six had a choice between one or no mobility devices, four had a choice between multiple mobility devices or no mobility devices, and four could not walk without mobility devices but had more than one mobility device they could use (e.g., walker or wheelchair). For example, one very small six year old explained that although she could walk, she used a stroller in crowded areas.

**Quote 18-2-11 (6 year old girl with a rare Neuromuscular Syndrome)**

**Child:** [points to a question in the green booklet]

**NY:** That one says you walked in crowded areas most of the time. And your Mommy wrote in for you that you have a student helper at school. What happens when you go to the mall?

**Child:** I go in a stroller.

**NY:** She puts you in a stroller?

**Child:** But I'm not a baby!

**NY:** She knows you're not a baby but is that easier for you?

**Child:** Yeah.

**NY:** Do you walk a little slower than mom does?

**Child:** Yeah. [points to a new question]

This was a costly trade-off for this girl, because although she gained efficiency to go shopping, she was often mistaken for a baby, and this was very detrimental to her self image. However, not all children experienced adverse consequences as a result of the use of assistive devices. A 15 year old boy with quadriplegia talked about how he always wore his legs and was always in his wheelchair outside the home (Quote 22-1-3). For him, his prosthetic legs and wheelchair were an integral part of his self image. However, apart from these two examples, the influence of self image was rarely mentioned.

Assistive devices were also chosen when they reduced the likelihood of adverse consequences related to activity, such as excessive energy expenditure, pain or other symptoms. This issue of adverse consequences was mentioned predominantly by children with arthritis. For example, a girl who wore special shoes to prevent foot injury and pain stated that they did achieve the desired reduction in pain, as compared to walking without them, but that the relief was only temporary. When asked about assistive devices she stated:
Among children with disorders other than arthritis, assistive devices were not frequently described as a method to avoid pain. These other children were more likely to choose not to do a painful activity. For these children assistive devices were more commonly a mechanism for increasing safety. For example, two of the children sometimes wore helmets for safety.

Quote 9-1-1 (9 year old boy with T12 spina bifida)
NY: What is one of the bad things about being in your wheelchair? You said sometimes it’s fun and sometimes it’s not.
Child: If I fall.
NY: Does that happen a lot?
Child: Once last year.
NY: And what happened when you fell?
Child: A friend, Justin, he pushed me and I banged my head.
NY: You hit your head.
Child: Yeah. I had my helmet.

Quote 12-1-1 (8 year old girl with diplegic cerebral palsy)
Child: I wear a helmet going up the stairs and outside at recess.

The latter child also talked about her walker being primarily to prevent falls.

Quote 12-1-4 (8 year old girl with diplegic cerebral palsy)
NY: Tell me what it’s like walking without your walker compared to with it.
Child: Easy.
NY: It’s easy without the walker?
Child: Yeah.
NY: So you use the walker why?
Child: Just in case I fall, but the walker makes it harder to fall.

A similar relationship between assistive devices and safety was expressed by another child:

Quote 19-3-5 (15 year old boy with congenital muscular dystrophy)
NY: So for a short distance you’ll go without your cane?
Child: Yeah.
NY: Will you do that even outside of your home? If you’re at school and you want to get up and get something from a friend’s desk, will you get up and go get it and come back or will you go for the cane?
Child: I think I would go for the cane because I’m in grade 10 and it’s an all boy’s school and they’re very rambunctious, and also the school is small and there’s 1200 people in the school and the thing is the school was made for only 600 so it’s over crowded.
Thus, assistive devices were described as the predominant method through which safety was enhanced for many children, and as a mechanism to prevent pain among children with arthritis. These effects were positive in nature, however, some negative effects (e.g., needing help with dressing because of splints) were also described. These examples also show that children again considered the three themes previously described in the trade-off for assistive devices and included several sub-themes or elements in the trade-off. These other elements included: the internal and external environment they were to perform in, the quality and efficiency of the performance, and the consequences in terms of symptom costs and safety risks.

**Modification of Technique**

Besides assistive devices, the children also described a variety of alternative techniques that they used to modify their approach to an activity to facilitate performance. Again, there was variability in the implementation of these modifications, that was related to the trading-off process.

One of the girls with arthritis talked about playing tag, and how she modified her speed of play to better suit her ability level.

**Quote 2-1-20 (10 year old girl with arthritis)**

NY: Do you think you can keep up with them now?

Child: Uh, for the first two minutes maybe. [Long pause] That’s like if I could jog it.

NY: Then what happens after two minutes?

Child: I slow down [emphasis on speed]; [if that does not work then] I don’t play anymore.

Other children talked about how they changed the biomechanics of the activity.

**Quote 3-1-18 (11 year old boy with arthritis)**

NY: What are some of your other tricks?

Child: Sometimes if I can’t like um [pause] wash my face, ‘cause after I, you soak it you have to turn it [motions wringing out the face cloth].

NY: You have to wring it out.

Child: Yeah. Sometimes I can’t do that, so I put my hand on the sink [motions bracing arm while he wrings the face cloth out] and I turn it so, and that gives me some help so I can, like, wring it out some more.

Another boy with slight disability related to spina bifida described his walking and stair climbing as having a “wobble.” He disliked the “wobble” because the other kids called him
names like “rubber legs or plastic legs.” He chose to modify his technique when he wanted to appear non-disabled.

Quote 6-1-13 (10 year old boy with L5/S1 spina bifida)
  Child: I have an easy way to go up the stairs.
  NY: Oh, tell me your trick.
  Child: Ok. Umm the stairs go up to about here on me [motions to just below his knee] and I lay my knee on it and I just go up. Every time I hit the step I lean on it
  NY: So then you go down on your knee?
  Child: No, like . . . [demonstrates walking up stairs bracing tibia against the step above].

Through this subtle change in technique he climbed stairs with a gait pattern that is difficult to discern as unusual in any way. Thus, self image (discussed previously as an adverse consequence of assistive devices) was a seldom mentioned but important element in the trading-off process, and was sometimes preserved by technique modifications.

**Active Use of Help**

Active use of help was the final way in which children modified their activities to facilitate performance. Consistent with their descriptions of using assistive devices and modifications of technique, their comments again showed variability in the active use of help, that was related to the trading-off process. The term ‘active’ is important to this element, since the children most often described accepting help for a component of an activity or as an assist. Passive use of help means that someone else completely did the activity for them. Thus, it is a form of opting out (not performing an activity) and does not represent a new factor in the process of determining performance.

The active use of help was most often accepted for activities for which the child had an affinity, but found formidable. There are many activities that children can do, but not in all situations and not necessarily without consequences. In these cases they often accepted help with the most formidable or time consuming portion of the activity. For example, one young girl talked about the cutting component of eating.

Quote 1-1-30 (10 year old girl with arthritis)
  NY: So how do you decide whether you cut it, or your mom cuts it or someone else does it for you?
  Child: Um [pause] usually my mom cuts it.
  NY: How come sometimes you cut it yourself? What happens on those times that you end up cutting it?
process. She also described the trade-off related to her stiffness symptoms (internal environmental element), and the trade-off made when mom’s help was not available. Thus, a similar pattern, to that of considering situational elements and consequences in decisions related to assistive devices, was seen regarding the active acceptance of help.

In another example, a 13 year old girl with arthrogryposis stated a clear preference for help with her hair when she was going to school.

Quote 24-1-1 (13 year old girl with arthrogryposis)

Child: All I need help with is getting stuff in my hair.
NY: Who brushes your hair?
Child: My mom usually in the morning.
NY: Do you think that's because she likes to do it or is there another reason?
[Mom enters and looks surprised]
NY: I guess we know the answer to that one. [It's not because mom likes to.]

Child: I tried doing it this morning and then I got my sister to brush it because it didn't turn out.

This girl explained the difficulty she has getting her arm and hand in position to brush her hair, and clearly the demands of the activity were very close to her maximum ability. However, part of the decision to have help with her hair was because she wanted to look good. This is an issue of quality of the performance, or the desirability of the outcome. Quality is a subjective judgment of acceptability, and for this girl, her own skill at hair dressing did not result in an acceptable end result. Thus, help was accepted.

The final important element, that came out primarily in relation to acceptance of help, was the desire for autonomy. Children frequently want to do things on their own.

Quote 2-1-10 (10 year old girl with arthritis)

Child: It’s not that I can’t. it’s just that; what did it say?
NY: You said [in the ASKp booklet] that you do it most of the times. I’m just curious who does it the rest of the time.

Child: My mom! When I’m going to school, I’ll be brushing my teeth or something and she will be combing my hair. It’s real annoying. It’s not that I can’t do it, it’s just that she will do it.
Another child, also with arthritis, had similar comments regarding help from his parents.

**Quote 3-1-27 (11 year old boy with arthritis)**

**NY:** Are there any times that you think your mom should help you with things, and you think no, no, no, I don't need help right now?

**Child:** Um [pause]

**NY:** Or do you always take her help.

**Child:** Not always. Sometimes she says "I'll help you with that" and I say "No I want to do it myself."

**NY:** How come you want to do it yourself?

**Child:** 'Cause I want to try, not everything I want my mom to do, 'cause I want to try and see if I can do it first.

These examples show how patterns of assistance develop around children with variable abilities and fluctuating symptoms. These patterns often develop for situational reasons (e.g., time constraints in the morning), but once established were difficult for the children to break free from. Thus, children have a built in care-giving system that is sometimes disabling in its efficiency. However, among the older children, the desire for autonomy appeared to be stronger and motivated them not to accept help. Among older children, assistive devices and modifications of techniques predominated, except where the discrepancy between ability and the demands of the activity was great.

In summary, the role of modifications in the trading-off process can be seen in the many examples in this section. Children most often modified formidable activities. Their accounts describe the use of a variety of assistive devices (including personal assistive devices, environmental modifications, and alternative modes of transportation), modifications in technique, and circumstances under which help was accepted. The modifications theme enters into the trading-off process along with affinity and formidable ability and is influenced by many of the elements previously discussed.

**Trading-off Process: Judgment of Acceptability**

The qualitative results so far have discussed the themes and elements that the children considered in determining their performance, but have not explicitly addressed the process. The questions of whether or not the children were aware of the process and how they integrated themes and elements in the process are important, and are addressed in this section.
It is important to recognize that although the themes and the concept of trading-off were common across children and across different activities, the process was not static. As demonstrated, different themes and different elements were considered in different situations. Thus, the relative contributions of different themes and elements to the outcome of performance appeared to be very fluid.

**Children’s Awareness of the Trading-Off Process**

Under the section of “how formidable the activity is perceived to be” the importance of the perceived discrepancy between a child’s ability and the demands of the activity with respect to performance was discussed. The magnitude of the ability-difficulty discrepancy also had an important effect on children’s abilities to articulate their trading-off process. Children were able to describe the factors determining the performance of activities close to their ability level much more easily and in greater detail, than those factors relating to activities further from their ability level. For example: when asked why they did not do something that was very difficult (relative to their ability level) they often said “because” or “I don’t know” and looked at me as if to suggest it was intuitively obvious. Apparently they had not done much thinking or trading-off for these extreme activities. For example, a seven year old girl who lacked lower extremity muscle control due to T12 Spina Bifida talked about why she did not dress herself. She loved the Power Rangers, so her Power Rangers clothes were used as a best-case scenario.

Quote 8-1-9 (7 year old girl with T12 Spina Bifida)
NY: Can you get into your Power Rangers stuff?
Child: Nope.
NY: How come?
Child: I don’t know.
NY: You don’t know? What would happen if you imagined trying to put your shirt on? Do you think you could put your shirt on?
Child: No.
NY: Why not?
Child: Cause it’s hard.
NY: It’s hard. What’s the hard part? [No answer] What’s the hardest part?
Child: Pants.
NY: Pants. Oh. What is so hard about pants?
Child: I don’t know.
It was very hard, particularly for younger children, to articulate their rationale for not doing something that they had never considered. A similar scenario was observed in a second young girl with the same condition, with respect to getting dressed in the morning.

Quote 7-2-12 (6 year old girl with T12 Spina Bifida)

Child: Um, my mom usually dresses me.
NY: On your bed?
Child: No. Downstairs.
NY: Downstairs?
Sister: I dress by myself.
NY: You dress by yourself? Well, that's good.
Child: No you don't (both laugh).
NY: Do you think you could put on those pants that you have on, on your own if you had to?
Child: No.
NY: No? So if mom was really busy, and said 'oh just try it.' Do you think you could ever manage to wiggle your way in?
Child: No. [Long pause]
NY: How come? [Very long pause] Don't know?
Child: I don't know about that one.

The parents of both of these children had routinely provided assistance, and this habit may never have been questioned or challenged.

When there was some possibility of achieving the activity, older children could articulate the limiting elements. For example, a 13 year old with arthrogryposis talked about how her balance and braces in skating prevented her from joining her friends in one of the group's favorite activities.

Quote 24-1-5 (13 year old girl with arthrogryposis)

NY: What's the hard thing about skating?
Child: I can barely stand on my own two feet. No, I can. It's just that it's hard with skates. I wear braces on my feet. You can't see them now 'cause I'm wearing socks, . . . so if I tie them [skates] up it has to be tight, but my braces wouldn't make them that tight, 'cause they won't go that tight. If that makes any sense.
NY: So you can't get your skates on as tight as you would want.
Child: Yeah. And I wouldn't be able to skate properly.

When asked about activities close to their ability level there was greater variability in the strategies used by children to complete the activity, and greater variability in the ultimate outcome. The themes and elements became increasingly important. Children explained in great detail the reasoning process, and appeared to be much more aware of their trading-off
dysophy process with respect to this group of activities. For example, one older boy with muscular dystrophy (MD) discussed in a general sense, how his trading-off process worked.

Quote 20-2-6 (15 year old boy with Duchenne muscular dystrophy)
NY: Are there other things that help you decide whether you're going to do something or not?
Child: If I feel up to it. Do I really want to do this? Um, am I strong enough to do this? I don't know. I think about that stuff to myself and see if I want, and I see if I can do it. If I want to do it I'll try, if I can't, I don't.
NY: Are there days that you're more motivated than other days?
Child: Usually. If it's a warm day and nice outside, I'll probably convince myself to go out and do something. I usually go outside with my friends anyway. Like even in this cold, like just go to the store and back to grab something to eat like cheese or stuff like that.

His account showed a great deal of insight into the themes of affinity and formidability, and also into the elements of internal environment (i.e., motivation) and external environment (i.e., weather). Another boy demonstrated the depth of his knowledge with regard to factors affecting the demands of getting in and out of the tub.

Quote 19-1-6 (15 year old boy with congenital muscular dystrophy)
Child: Yeah. That [meaning tub transfers] I need help with because, because um, without, because of the support and balance.
NY: So what would happen if you did it by yourself? If you decided that mom's busy, I'm not going to bother her? What would happen?
Child: Um, it would be really quite hard to actually get in. To actually, like because without shoes on and that sort of support I have not actually walked a lot barefoot. I still have scars on my foot from the surgery still. I still have that and its sort of walking there, and also there's nothing really to hold on to, you know what I mean, and that's a problem. I could probably slide myself in, but then it's hard trying to get myself back out.

These two examples came from boys of the same age and diagnostic group, but with different cognitive levels and vastly different degrees of physical disability. The boy with Duchenne MD was a full-time wheelchair user and transferred by Hoyer lift. The boy with congenital MD was a full-time ambulator with a wheeled walker who transferred independently. Both were able to clearly articulate a trade-off for activities close to their ability levels. Children with muscular dystrophy may be more cognizant of the factors and trade-offs than other children, because of the progressive nature of MD. However, these examples were not unique to muscular dystrophy.

Many of the younger children also identified many elements on which the outcome of activities close to their ability level depended. This was achieved through the use of examples,
because these children were less cognizant of their trading-off process. The girl who gave the following account could not summarize the elements as succinctly as the older boys had, but could establish some associations between symptom levels through the day (internal environmental element) and performance of activities through sequential questioning.

Quote 1-1-18 (10 year old girl with arthritis)
NY: Are there good times to do up your shirt, or are there good shirts?
Child: Um, sometimes I can get them up.
NY: Ok, what’s different about the times when you can get them up? [Pause] When can you get them up?
Child: I guess when um, um I can’t get them up when I’m stiffer. And I can get them up when I’m not as stiff.
NY: Is it easier to do them up, or undo them? Or does it matter?
Child: Well its sort of easier to undo them.
NY: It’s easier to undo them. Do you think it has anything to do with when you’re doing it? [Long pause] The time of day that you do it?
Child: Well, um, yeah, sort of.
NY: What’s a good time?
Child: A good time, is I guess, in the afternoon.

Through an analysis of such accounts many elements, that contributed to how challenging the activity was for them, were revealed. Many of these elements related to the situation or environment, while fewer related to the consequences of the activity (e.g., potential for injuring themselves). Thus, the children were aware of the themes and elements in the trading-off process, and were able to articulate them.

Integration of Themes and Elements

Having identified the structure of the trading-off process (the three themes and five elements), what remains to be defined is the process through which these components were integrated.

As was alluded to in the example of the 13 year old girl with arthrogryposis accepting help to style her hair in the morning, there are usually judgments made about the acceptability of a particular approach to an activity or the expected results. These judgments are a final qualitative assessment (act as a gate keeper, or final stamp of approval), for any performance being considered in the trade-off. Just as many children refuse to wear glasses in front of their
peers, whether or not a choice of physical behaviour was acceptable to the child had great
influence. For example, one child talked about her desire for having her shoes tied tightly.

Quote 13-2-2 (13 year old girl with diplegic cerebral palsy)
Child: No, I never do my shoes.
NY: Do you think you could do your shoes?
Child: Well, it matters how hard they are kind of thing. If they’re Velcro, sure. Tying them up, yeah I
can do it but it matters how tight. I can’t get it tight, that’s the only thing. I won’t have enough
support. It won’t be strong enough [so mom ties them].

In this example her desire for tight shoes and stable feet was an expression of disability, in
that her shoes needed to be tied tightly because of inadequate muscular support. She traded-
away independent performance and allowed her mother to tie her shoes in favor of stable feet.
She not only considered all the factors, but she assigned greater importance or value to having
her shoes tight, than to her autonomy. In this situation her final judgment was to accept help.

In the trading-off process the children made a final judgment of whether performance of an
activity was worth it. Since all the judgments made about performance required a subjective
comparison, the interviews were first analyzed to determine the standard to which the children
were making comparisons. Most often it was a normative standard either observed in other
children (e.g., siblings), or experienced by themselves at a different point in their life (e.g.,
compared to ‘good days’). For example, one girl with arthritis, who spoke at length about
things that were hard for her, stated that the same things were “markedly cinchy” for her twin
sister (Quote 5-3-3). Others compared themselves to teachers’ expectations for the class as a
whole. For example, one girl with arthritis talked about not being able to meet the teacher’s
expectation of finishing the one mile run in gym class at the same time as the rest of her class
(Quote 2-1-2), and another talked about the difficulty she had keeping up with her class when
copying copious notes from the blackboard before the teacher erased them (Quote 5-3-6). These
examples may also be considered a normative standard.

Children from all diagnostic groups, both boys and girls, articulated these judgments
across most activities. However, direct references to a judgment were less prevalent among
younger children. This is likely because the process of making an abstract comparison is
beyond the cognitive abilities of younger children. Furthermore, the judgment of what was acceptable varied in different situations because they may use different criteria. Thus, their judgments reflected some fluidity over time and circumstances.

The children’s judgments also reflected concern about style. Style was not discussed by many children, but two in particular emphasized its importance to them. One example was a 10 year old boy who had the choice of walking without braces, but chose to wear braces because the style without braces was unacceptable to him.

Quote 10-2-6 (10 year old boy with L4 spina bifida)
NY: What’s the problem when you walk without your braces [reported in booklet]?
Child: It’s wobbly... I already showed you, my feet bend in like this.

This boy’s sub-talar joint lacked stability (impairment) and caused excessive lateral weight shifting when walking without his braces. This style of walking was unacceptable to him, and he traded wearing braces in favor of a walking style more consistent with that of his peers.

Another element considered in the judgment process was the relative efficiency of doing the activity by one method versus another (e.g., with or without assistive devices). Simply stated, children often found it quicker and easier and therefore more efficient, to use a particular assistive device for a particular activity. Children talked about efficiency in terms of ‘ease.’ When probed further this ‘ease’ seemed to translate to amount of energy expended. This was confirmed in later interviews.

Quote 22-1-3 (15 year old boy with quadrimelia)
Child: Well when I was younger I used to walk all the time, but now I don’t. It’s easier in my chair if I can move around.
NY: Does it tire you out to walk around?
Child: On my legs? Yeah.
NY: Do you walk at all now to get into the bathroom or small places; will you choose to walk?
Child: No.
NY: You like the chair?
Child: Yeah.

Each of the themes and elements previously introduced entered into the trading-off process, and was reflected in the final judgment and subsequent performance of activities. What has been evident in the examples, but not articulated, is that these factors are not equally considered (or weighted) in the trade-off process. When style was considered, the children
sometimes valued style more heavily than other elements. At other times, they valued

efficiency more heavily. Thus, the outcome of the trade-off was dependent upon which aspect
(style or efficiency) was more heavily valued in the trading-off process for a particular activity.

For example, use of a wheelchair clearly did not approximate non-disabled walking in terms of
technique, but did approximate the non-disabled standard in terms of speed. For this girl,
speed was the heavily valued component of mobility.

Quote 13-2-7 (13 year old girl with diplegic cerebral palsy)
NY:  When can't you keep up with them when they're inside?
Child: When they go too fast.
NY:  If they walk fast, can you keep up with walking fast?
Child: Yes. Running fast you've got to be kidding.
NY:  What about in your chair, you must be fast in your chair.
Child: Well yeah, in my chair, but when I walk nah.
NY:  Are you faster with your walker or without it?
Child: Well, I'm faster with my walker. I'm faster [still] with my chair.

This child also chose a wheelchair for outdoors and for longer distance activities. Thus, it was
not only presence or absence of the multiple themes and elements in the trade-off that must be
considered, but also how much the child values each theme or element. In this case, speed
(efficiency) was valued over style.

Thus far three themes (affinity, formidability, and modifications) and multiple elements
that contribute to an extensive and intertwined trading-off process have been described.
However, there was one final issue that children sometimes were forced to considered in the
trade-off: exceptions. Exceptions were most often made to stay within the rules set by
parents, teachers, or clinicians. They have not been included as a theme or element because
they did not show the consistency across individuals and activities that the other themes and
elements did, but were included here because exceptions have the potential to override the
entire trading-off process. For example, one girl talked about how she decided whether or not
to participate in jazz. She was extremely articulate about elements that influenced her
performance, but also articulated very clearly that the rules dominated at times.

Quote 5-1-16 (11 year old girl with arthritis)
NY:  You mentioned that you like jazz. You wore your jazz costume for Halloween.
Child: I'm, I can't stay in the splits for long... It's sore though. It hurts a lot here [adductors]. Even
though I do the warm-up.
This example shows that although children have a very intricate trading-off process that controls their performance of activities, the entire process may be circumvented by rules. There may have been occasions on which the children entered the rule into the trade-off process and may have decided to risk the consequences of disobeying rules to complete the activity. However, because rules were rarely discussed by the children, detailed information was not available.

Vignettes

The final issue in the analysis was whether these themes and process were expressed across individuals, across activities, across age groups, and over time. The previous quotations have represented all the interview subjects and most of the activities. Although the interviews were conducted during a five month period, several of the subjects reflected back over their life span. Two examples are used here to demonstrate the consistency of themes and trading-off process across activities and over time. For descriptive purposes they have been named Tammy and Timmy (akin to Easter Seals’ disability representatives).
The first example was a 15 year old girl with congenital quadrimelia, nicknamed Tammy. Although she lacked legs, in the usual sense, she did have very short legs with feet and toes attached to her hips. The one aspect of physical disability that was most pervasive in Tammy’s accounts was her absence of hands. She discussed in great detail the trade-offs she made in different activities with respect to arm and leg prosthetics. The pattern of Tammy’s trade-offs indicated that efficient function and quality of performance were most important to her.

As a child Tammy was fitted with both arm and leg prosthetics, but found they interfered with the use of her toes.

Tammy: I used to always do everything with my feet.
NY: And now you're using your arms more?
Tammy: Yes, I can't use my feet now at school because my legs are on, and I can't sit on the desk because I'm in high school. But in like public school I sat on the desk and I could use my feet, in the younger grades.
NY: Do you remember what grade you were in when you stopped?
Tammy: I think in the younger grades. Some of the grades I wore my legs and not my arms and in some of them I wore both and in some of them I wore just my arms. When I was in kindergarten I remember I wore both of them. Grade one I don't remember and in grade two I wore just my arms I think. It was always different.
NY: How did you decide to change from wearing both of your arms and legs to just wanting to wear your arms?
Tammy: I don't think it was my choice then, I don't remember, I was only little. But now like at school I only wear my legs. If I wore my arms, I wouldn't be as independent. If I wear both, I can't do all the stuff.
NY: A lot of kids say that about the prosthetic arms that they do much better without them than with them.
Tammy: Yeah.
NY: What about your legs, do they give you an advantage?
Tammy: Yeah 'cause I'm taller and I can probably walk faster. I don't know. And I can reach more.
NY: But you don't tend to wear them at home or do you?
Tammy: Not really, I don't need to.
NY: What's the difference at home that you don't need to?
Tammy: At school I really really wear it to be tall. Because if I was walking around school little I'd get trampled.
NY: So part of it is a safety issue then.
Tammy: I can also wear better clothes if I want to ... But here I really, I don't need my legs.
NY: So if you were to figure out how much your decision to wear your legs had to do with looking better and being able to wear nice clothes, is that a big part of it or just a bonus of wearing them?
Tammy: Um.
NY: Is that part of what makes you put them on or ...?
Tammy: I don't know cause I can probably get the same clothes except then I can wear jeans but now I can wear jean shorts. [Pause] I don't know. It looks better too when I'm taller.

Later on in the discussion she clarified some of the aspects of the external environment that are responsible for her trade-off in favor of lower limb prostheses.
Tammy: I wear my legs . . . to go out.
NY: So anytime you venture outside of the home?
Tammy: It depends where I’m going. If I was going to the park or something, or I’m at camp or something, I wouldn’t. If I was going to the mall or out for dinner then I would.

Because prostheses use was optional for Tammy, she provided vivid information about the trade-off process. With her prostheses she was approximately five feet tall, could wear standard clothing, and appeared non-disabled. Without her prostheses she was under three feet tall, was restricted to shorts, and was obviously different anatomically from those around her. The prostheses improved her physical appearance (style) in all situations, but the quality of her other activities was adversely affected by her choice. In the park or at camp the activities of interest were not only walking, but playing. Tammy’s style of play was adversely affected by the prostheses. However, going to the mall or to dinner had fewer demands for functionality, and her preference was for a stylish appearance. Thus, Tammy favored assistive devices when they improved the style of her performance relative to the standard without any adverse effects.

Tammy also articulated her trade-offs with respect to assistance.
NY: Now, you comb your hair once in a while [according to her booklet]?
Tammy: That’s different. I can comb my hair all of the time now, but when I go to school, my sister does my bangs for me ‘cause I can’t do that. Because my hair will look ugly. I can brush it though. During the summer I’ll brush my hair by myself every day.

Thus, quality of appearance is important enough to permit the acceptance of assistance when a social situation is involved (school).

Tammy also articulated her trade-offs with respect to mobility devices. Although she used prosthetic legs for dining out, she would trade them for a wheelchair when the activity was visiting a theme park.

Tammy: If I was to go to Wonderland, I’d bring the chair. But it’s not that I wouldn’t be able to do it [meaning walk at Wonderland with prostheses], I would just be slow. But I could do it, it’s just that I’d be slow.
NY: So you take a manual chair and somebody pushes you?
Tammy: It depends where I am. I could take my scooter.

Although the mode of transportation that provided the greatest efficiency and independence simultaneously was a scooter, she explained that to take her scooter to Wonderland required a special van for transportation to Wonderland. Thus, although it was most efficient at
Wonderland, it was less efficient when the whole process of the Wonderland visit was considered. In this set of circumstances Tammy traded away independence and quality in favor of an efficient means of getting around the theme park. The option of staying home was not mentioned.

Tammy was a member of a large and very supportive family. Their behaviour with respect to their daughter did not in any way suggest that they viewed her as 'disabled' in the way that many children become labeled by the medical system. Rather, they encouraged her independence, and permitted her to make her own decisions and trade-offs. In discussions with Tammy's mother, it was clear that in their household capability was only half the picture. What her children did with their abilities was more important.

The second example was a 12 year old boy with arthritis, nicknamed Timmy. Unlike Tammy whose condition was congenital and therefore chronic, Timmy's condition was diagnosed when he was 11, and had responded well to treatment over the 10 months prior to the interview. At the time of the study, Timmy experienced mild disability, detected only by the ASKp, not by either the ASKc or CHAQ. However, Timmy's accounts reflected back to before his arthritis, during the worst phase of the disease (10 months earlier), and to more recent (mildly disabled) times. The one aspect of physical disability that was most pervasive for Timmy was pain. He discussed in great detail the trade-offs he had made primarily in athletic activities. The pattern of Timmy's trade-offs indicated that the continuation of non-disabled activities, but with avoidance of pain, was most important to him.

Timmy described the first time that he remembered accepting help because of his arthritis.

NY: Was there ever a time when your arthritis was worse that they did things for you?
Timmy: Yeah. When I was walking to my cottage in the snow, my leg was like really, really sore and my dad carried me up to the cottage.
NY: Were you glad that your dad carried you?
Timmy: Yeah [pause].
NY: What else did you think about when your dad had to carry you? What did it feel like for you?
Timmy: Mostly being hard for him.
NY: You were worried about your dad?
Timmy: Uh-hmm.
NY: Do you think walking in the snow is going to be tough again this year?
Timmy: It depends if there is lots of snow or not.
Timmy was aware of some of the elements (e.g., formidability) that cause his pain to worsen, and his performance to deteriorate. He was also aware that it may be a burden for his father (a fairly healthy man in his early 40’s) to carry him. It is unclear whether or not his own autonomy contributed to his trading-off in favor of being independent until assistance was offered. However, it was clear that his symptoms must have been severe enough to alter his performance (i.e., make him limp) to elicit the offer of assistance in the first place.

Timmy also discussed the effect the medication had on his symptoms.

His attribution of pain relief to the medication and assistive devices explains why he continued to be compliant (trade-off in favor of the medication) even where there are significant side effects (consequences).

NY: What was hard about brushing your teeth.
Timmy: ‘Cause I get these sores in my mouth from the medication I’m taking.
NY: Oh so the medicine gives you sores inside your mouth.
Timmy: Um hmm.
NY: And do they hurt when you brush your teeth?
Timmy: Yeah, sometimes. Sometimes I can’t eat [pause] because it hurts [mumbling badly].
NY: Because of the pain?
Timmy: Um hmm. I still have one like right here in my mouth and it really hurts. It’s probably the worst one I’ve had.
It is apparent that he was aware of the components of the movement that were most painful. Throwing with his left was not judged good enough, so he elected to trade his usual position on the team for the position of pitcher to avoid pain. However, at other times his motivation and enjoyment of an activity appear to be sufficient to overcome the consequences.

Timmy mentioned that he was now playing trumpet. This is an activity that requires rapid finger movements in a sustained gripping position of the hand, and is therefore very likely to cause pain in anyone with arthritis in their hands.

Child: I'm good at the trumpet. I just started it in the beginning of September.
NY: Do you want to play in the band?
Child: Yeah. I'm in the grade seven band.
is one method of supporting the generalizability of the results to the larger study population. The consistency of concepts across groups (shown through groups) and examples of the same concepts across activities and time groups (shown through groups) and examples of the same concepts across individuals of different gender, age and diagnostic performance have explained the generalizability differences. Consistency of the concepts of reading-reading of tales explains the generalizability differences. The qualitative research has evidenced these findings by demonstrating the process of reading that explains the generalizability differences. The hypotheses support the consistency of the qualitative differences reported between capability and performance in this process. Differences between capability and performance (Anderson et al. 1977, Yoon 1994, Yoon et al. 1994) and differences among children with disabilities come from previous qualitative research showing the importance of examining the factors that may contribute to variable performance in the process of reading.

Generalizability of Training-On Process

The process across a range of activities and situations within each individual activities for both children. Thus, these findings demonstrate the consistency of the training-on process. These two children of different ages, with different disorders, from different families, and with different disabilities, and with different experiences, As illustrated with the other two children, depict two children of different ages, with different disorders, from different families, and with different experiences. Where Lanny had the advantage of learning from her physically disabled siblings, Tommy bad uncertainty as to what levels of activities were advisable, but he was generally supportive. Tommy’s home was quite different from Lanny’s. He was one of three children, and the social environment and medication combined made it worth it for him to play an instrument.
Support for the generalizability of the concept of trading-off and its associated themes beyond this study population must come from the literature [Conrad 1990]. Five qualitative studies were identified in the literature that have direct bearing on the concepts raised in the present research. This concept and its themes are evident in other literature, across different populations and extend beyond the realm of health research.

Strauss et al.'s 1984 text describes a process similar to trading-off [Strauss et al. 1984]. Their work is based on a variety of authors' qualitative work, and represents an overview of many issues pertinent to the quality of life among those experiencing chronic illness. They describe the process as one of “juggling” competing influences (both positive and negative). This process is described with respect to preventing crises, managing care routines, controlling symptoms and planning one's day. Their use of the term “juggling” is beneficial in that it suggests a very fluid process, just as the trading-off process proposed here is described as being in flux. However, “juggling” also implies an attempt to keep all the factors in the air at one time. This was clearly not the intent of many of the children interviewed in this thesis, who were aware that some factors had to be dropped in favor of others, because they could not balance them all.

Strauss et al. [1984] make the interaction of factors in the decision process explicitly clear with respect to compliance with medical treatment routines. In this context they suggest that compliance is associated with: trust in the prescribing clinician, competing recommendations, evidence of effectiveness, side effects, interference with important daily activities (both their own activities and others’ activities), and the perception that the positive aspects of treatment are not “outweighed” by a negative impact on their sense of identity. Although the focus of the text is on quality of life, rather than performance of activities, many similarities are obvious between Strauss et al.’s factors, and those identified in the current research. For example, side effects are consistent with the children’s description of the role of internal factors (e.g., symptoms). Strauss et al. [1984] also describe “normalizing tactics” that are used to hide visible evidence of chronic illness in an attempt to “pass” for normal. Their examples are
particularly consistent with the account of the 10 year old boy with L5/S1 spina bifida, who described a modification of stair climbing in an attempt to pass for normal among his peers.

Locker and Kaufert [1988] describe a “trading off” process among 10 polio survivors for the purpose of enhancing quality of life. They describe these individuals’ considerations of the positive effects of accepting mechanical ventilation support (i.e., gains in energy, mobility and well being) as being balanced against the negative effects (i.e., altered routines, new worries, and loss of self-sufficiency). Thus, they identified many concepts similar to those generated by children in the present research. For example, the polio survivors whose demands of activities closely approximated their respiratory capacity (termed formidable in this thesis) described a more extensive use of alternative coping strategies (e.g., “pacing” and “short cutting”), where this thesis reported a similar behaviour for activities close to ones capability. The gains in quality of life by post polio survivors were influenced by the degree of disability, but also by the type of technology available to them (similar to ‘availability of modifications’). Issues of time and scheduling (akin to ‘external environment’ factors), consequences of performance, autonomy and efficiency were also common to both analyses.

Weiner [1984] discusses the trading-off process as a balancing of limited options in her study of 21 adults with rheumatoid arthritis. Specifically, she describes a “race” between the physiology of the individual with rheumatoid arthritis and the activity demands of the real world. In this “race” one “imperative” (or factor) would overtake the other at any moment, only to later be overtaken itself (switching between physiology and activity winning). This paper also discussed the impact of internal environmental factors (e.g., symptoms) and external environmental factors (e.g., being at home vs. out in public) on performance of activities. There was particular emphasis on the role of consequences of performance on the trade-off, with minimal emphasis on autonomy and quality of the performance. This population was unique in that uncertainty played a much larger role than in the sample of children with disability, but the predominant concept was a balancing of limited options or trading-off.
The concept of trading-off is also evident in literature outside of chronic illness. Backett et al. [1994] combined observational and interview data from three studies in Wales and Scotland. The focus of their qualitative analysis was to determine through what mechanism healthy lifestyle behaviours became integrated into behaviour. They reported that the translation of advice (related to capability) into behaviour (performance) was based on a "trading-off" of positive and negative aspects of the behaviour, and that this process was widespread across behaviours related to diet, exercise, alcohol and smoking.

Launer [1993] interviewed 36 new mothers in Indonesia for the purpose of determining to what extent and through what mechanism they adapted their work to accommodate breast-feeding of their infants during the first year of life. This paper reported a number of modifications (e.g., creative employment schedules), most common at the 13 to 31 week period. They identified predominantly cultural factors (e.g., the societies support for maternal proximity in the early neonatal period) and external resources (e.g., the number of other workers in the family and seasonal variations in the amount of work) as the predominant factors women integrated into the "trade-off" made between work and the demands of breast-feeding. This evidence supports the concept of trading-off. Evidence for the themes, and particularly the elements was not found here, however, this degree of detail and similarity was not anticipated.

Finally, Law's participatory research with 22 families of children with disabilities describes an ecological or interactional model of physical disability among children three to 12 years of age [Law 1992]. She describes how "inaccessible built environments" and an "emphasis on normality" contribute to the expression of physical disability in children. Although this thesis focused on the experiences of families, in particular the social implications, a model of childhood physical disability was included. In this model, shown in Figure 9, Law includes "environmental supports and barriers", "activity demands" and "child's skills" as factors contributing to "activity competence and satisfaction". These factors are very similar to the themes of formidable and availability of modifications presented in this thesis.
Thus, Law's research with parents supports the generalizability of this current research with children.

Figure 9  Law's Model of Childhood Disability


SUMMARY OF EXPLORATION OF DETERMINANTS RESULTS

Twenty-four children with a wide range of physical disabilities were interviewed to explore factors responsible for the difference between capability and performance. They described three main themes and five elements that determined their performance for activities that they had the capability to do. These themes were affinity for the activity, how formidable the activity was perceived to be, and the availability of modifications. The elements (or sub-themes) were: internal environment, external environment, quality and efficiency of the performance, desire for autonomy, and consequences of the performance.

These themes and elements were integrated by a detailed trading-off process. In the trading-off process described here, the children considered the value of each theme and
element, and made a final subjective judgment as to whether or not the end result would be worth it, given the situation. The children’s performances of activities were a direct reflection of this final judgment. If they had no affinity for an activity, it was not performed. If they had some affinity, but the demands of the activity were well within the child’s ability (activity was not formidable) it was performed by the usual non-disabled method. If the activity was formidable it was either performed by the usual non-disabled method (and the consequences paid), modified (by the use of assistive devices, change in technique, use of alternative transportation modes or active use of help), or not performed (because of failure to find, accept, or carry out an appropriate strategy).

The main concept of trading-off is supported by the literature. The same literature also lends support to the three main themes, and to a lesser extent lends some support to the five elements. As previously emphasized, the trading-off process was not a static process. There was a fluidity to this process as demonstrated in the detailed vignettes. Yoshida has previously described a similar fluidity in her qualitative analysis of the “reconstruction of self” among spinal cord injured adults [Yoshida 1993]. In her analysis the reconstruction is termed “pendular” to reflect this fluidity demonstrated by a shifting of spinal cord injured patients’ identities (e.g., from the identity of their former non-disabled self to the disabled identity as total self) over different points in their post-rehabilitation life. The shift covered both extremes, from an identity of very able to very disabled. The implication of fluidity in this thesis, is that knowing the themes and elements offers a significant improvement in estimating the performance of any child, but does not predict performance with certainty. Further research needs to be done in this area.

There was one caveat to the trading-off process described here. If a child was told by a significantly influential person that they must perform an activity in a specified way (e.g., use a brace) to prevent future problems (e.g., to preserve joint function), their behaviour did not necessarily follow the trading-off process. In this situation, current preferences and judgments may be traded-off in favor of potential future benefits.
Thus, through a complicated trading-off process, children determined how their disability was expressed as performance. The three themes and five elements identified as influencing performance were also likely responsible, in part, for the quantitative discrepancy between performance and capability for two reasons. First, because these interviews were restricted to accounts of activities that children were capable of completing but did not always perform. Second, because the quantitative analysis demonstrated that performance scores were much more strongly associated with the capability-performance discrepancy than were capability scores.

Table 21 Summary of Results Part B

<table>
<thead>
<tr>
<th>Primary Objective</th>
<th>Purpose</th>
<th>Results</th>
</tr>
</thead>
</table>
| Exploring Determinants of Physical Disability | 7 Difference between Performance and Capability: magnitude and statistical significance | n=200 (n=200 proposed)  
- ASKc scores statistically exceeded ASKp scores (p<0.0001), but magnitude depended upon aggregation method |
| | 8 Factors that Mediate between performance and capability: nature of the difference | n=24 (n=40 proposed)  
- children determined their performance by trading-off. The 3 main components considered were affinity for the activity, formidability of the activity, and availability of modifications/resources. |
CHAPTER 6

DISCUSSION & CONCLUSIONS

The present research has addressed the measurement properties of the ASK and the impact of contextual factors on the expression of disability as two separate issues. However, it is through the intertwining of measurement and conceptual results that the present research enriches our understanding of disability and enriches the ASK in the process. The following issues will be addressed in this chapter: the need for outcome measures, the advantages of a multi-method approach to scale development, the ASK's measurement properties, the implications of measuring capability versus performance, the implications for the conceptualization and measurement of disability (with respect to paediatrics and the broader implications for measurement in general), an integration of the multi-method approach, the generalizability of results and limitations, and future research directions.

Need For Outcome Measures

The review of literature explained that reduction in activity restrictions is a common objective of therapy in children with musculoskeletal disorders, and that disability measures are the most appropriate outcomes for clinical research in this population. The dearth of research evaluating the effectiveness of therapy for children, in terms of physical disability, was attributed primarily to the absence of appropriate measures. The review of literature also discussed the problems with existing measures (either disease specific, have not been adequately tested, or are not child-centred). Thus, there was adequate justification for the development of a child-centred disability measure for children (the ASK), and the continued research into its conceptual basis and measurement properties that has been the focus of this thesis.
Multi-Method Approach to Scale Development and Assessment

The ASK was developed by primarily clinimetric methods. The intent was to use it to assess and monitor the physical component of disability among children with musculoskeletal disorders. The conceptual framework of the ASK was tested from a variety of perspectives, using eight sub-studies that drew upon four different methodologic approaches. The approaches were: item characteristic analysis (assessment of item response patterns by Rasch modeling), clinimetric evaluation (comparison to clinicians definitions of disability), psychometric evaluation (comparison to other measures' operational definitions of disability), and qualitative analysis of disabling factors (analysis of children’s accounts of disability).

These four approaches may be considered methods of validation, since confirmation that a scale measures the construct it was intended to measure is an assessment of validity [Last 1995]. The four methods used, and the information they provided on the conceptual basis of the ASK, are discussed in the following sections.

First, item characteristic analysis (Rasch modeling) was used to assess the ASK's conceptual foundation by assessing the unidimensionality of the scale. This assessment (sub-study one) identified a single stable model of item difficulty patterns. Because the model converged and because no items had excessive misfit statistic, the analysis confirmed the overall unidimensionality of the model. However, there were a few items that did not fit this model as well as the majority of items. This misfit suggested that the few items in question may not belong to the same domain as the majority of the items in the ASK. Appendix III recommends the removal of 23 items to further refine the item content and more closely represent a single construct (i.e., physical disability).

Although the Rasch model was able to specify that the ASK measured a single construct or dimension and identified those items that did not represent the construct well, the model is a probabilistic mathematical model. As such, it has no power to identify what the single construct represents (e.g., it could be autonomy or anxiety rather than physical disability).
Evidence that the single dimension was physical disability came from the ASK development research, during which a focus group was held. Using the ICIDH as their conceptual model, this focus group selected the initial item content of the ASK and confirmed that the construct measured by the ASK and identified by the Rasch model was physical disability [Young 1994]. The addition of clinimetric, psychometric and qualitative methods in this thesis lend support to the conceptualization of the ASK as a measure of childhood physical disability.

Secondly, a clinimetric approach to validity was taken. The ASK was compared to two different clinical standards (global assessment of physical disability, and structured clinical observation) to ascertain whether the ASK was consistent with clinicians’ conceptualizations of physical disability. The strong correspondence with clinical assessments confirmed the conceptual basis of the ASK as a measure of physical disability, and thus its validity from a clinimetric perspective.

Thirdly, a psychometric approach to validity was taken. The ASK was compared to two other metrics (the CHAQ and HUI3) to ascertain whether the ASK was consistent with other scales’ constructions of paediatric physical disability. The strong correspondence with other metrics further confirmed the conceptual basis of the ASK as residing within the domain of physical disability and its validity from another alternative perspective; the psychometric perspective.

Fourth, a qualitative approach was taken. Interviews were conducted with 24 children to learn about the distinction between the ASKp and ASKc from the children’s perspectives. The terms used and explanations given by children in distinguishing between these two versions of the ASK confirmed that both versions measured physical disability, thus supported previous focus group research [Young 1994]. The children’s accounts suggested the items were all activities limited by physical impairments rather than by choice alone, and as such belonged to the construct of physical disability. Thus, the qualitative analysis further confirmed the conceptual basis of the ASK; this time directly from the perspectives of children.
In summary, all four methodologic approaches converged to confirm the conceptual foundation of the ASK. The item characteristics confirmed unidimensionality. Interviews with children and an inter-disciplinary focus group of clinicians during the ASK development research [Young 1994] support the assertion that this single domain was and is physical disability. The evidence on validity with respect to clinician’s assessments, clinician’s global ratings, the CHAQ and HUI3 all confirm the ASK remains a unidimensional measure of physical disability. The complete convergence of results from multiple methods and perspectives reinforces the strength of the conceptual foundation of the ASK. This convergence also highlights the benefits of the multiple inter-related sub-studies in this thesis and of linkages to previous research.

ASK Measurement Properties

Through the multi-method approach the measurement properties of the ASK were also documented. The purpose of the ASK was to provide a method with which to evaluate the effectiveness of therapeutic interventions at controlling physical disability. Thus, its purpose was to monitor change in physical disability [Kane and Kane 1981]. However, because the purposes of measurement are hierarchical [Kane and Kane 1981], it was expected that if the ASK was capable of monitoring, it would also have sufficient measurement properties to serve the purposes of description, screening and assessment. This thesis yielded evidence to support the ASK’s application for multiple purposes (description, screening, assessment and monitoring).

The measurement properties necessary for description were met by documenting face validity during the ASK development research [Young 1994]. High internal consistency was demonstrated in the present research. Construct validity and ability to detect disability in large groups, necessary for screening, were documented in sub-studies four and five by comparing the ASK to the CHAQ and HUI3. The ASK was found to be related to these other measures, yet had no ceiling and floor effects. Ceiling effects were found for both the HUI3 and the
CHAQ. Furthermore, the ASK had no problems with floor effects that were apparent on the alternative measures. Moreover, the ASK could be administered by mail, adding to its feasibility in large screening studies. Criterion validity, necessary for assessment, was documented in sub-studies two and three in which the ASK was shown to be concordant with both clinicians' global reports and observation. Furthermore, the distribution of ASK scores was negatively skewed, matching the distribution reported by the referring clinicians and that reported in previous population studies [Hanvey et al. 1994a]. The CHAQ's distribution was rectangular (inconsistent with other indicators of disability). Test-retest reliability and responsive, necessary for monitoring, were documented in previous research that showed excellent reliability [Young et al. 1995] and in sub-study six that addressed responsiveness. The ASK detected large effect sizes in the sub-group tested, and was at least as responsive as the CHAQ. Thus, the ASK fulfilled the expectations for a monitoring measure.

The assessment of the ASK's measurement properties went one step further. Whereas Kane and Kane's purposes and Feinstein's criteria for measures focus on characteristics of summary scores, the properties of individual items are also important to the optimal functioning of the overall scale. The appendices have shown the validity and responsiveness data for the individual items. In addition, item characteristic analysis (Rasch modeling) was applied to the ASK to test the characteristics of the individual items. This analysis confirmed validity of the items through fit statistics, and described their location according to the level of disability they measured. This depth of item analysis has previously been performed for only two paediatric disability measures: the PEDI and WeeFIM (but found two dimensions within the WeeFIM). Thus, the item characteristic analysis added to the conceptual understanding of the ASK. It also provided the basis for recommending that the ASK be reduced to 30 items from 53 items (see Appendix III). Analysis of the shorter item set confirmed this was possible without loss of information. This item refinement offers a significant reduction in respondent burden and makes the ASK comparable in length to the CHAQ, HUB, JAFAR and WeeFIM, and shorter than the PEDI and GMFM.
Issue of Capability vs. Performance

Having documented the strong measurement properties of the ASK, it was then used to address a conceptual question. The question was whether or not capability was different than performance and why. The quantitative comparison of ASKp and ASKc scores documented that capability consistently exceeded performance, and that the discrepancy was slightly greater among children with more severe disabilities. Because there was no existing theoretical explanation for this difference, qualitative methods were applied to generate an explanation. The explanation has implications for the use of the ASK versions, but more importantly, it has implications for the conceptualization of disability.

The qualitative analysis identified that the two versions of the ASK were related through a trading-off process. This process was common among the 24 children with physical disabilities who were interviewed in the present research. The results suggested that these children considered three predominant themes in determining their performance of all 53 activities. These themes were: their affinity for the activity, the formidability of the activity, and the availability of modifications and or assistive devices with which to complete the activity. In addition, five elements were identified that influenced the expression of disability across many (but not necessarily all) activities. The elements included: aspects of the internal environment (person), aspects of the external environment, quality and efficiency of the performance, desire for autonomy, and consequences of performance.

Many of the elements that were identified by the children as altering performance (i.e., ASKp scores) may be considered either social or environmental. Thus, the results of the qualitative research component (sub-study eight) have two possible interpretations.

The first interpretation is that the situational effects render part or all of the ASKp in the domain of handicap. Thus, the ASKp is not restricted to the domain of disability. This interpretation is consistent with the WHO model, because it distinctly includes environmental factors as an influence on handicap [WHO 1980]. If this interpretation is accepted, the results
from the comparison of capability and performance may be seen as a challenge to the conceptual foundation of the ASKp as presented so far (i.e., as a measure of physical disability). However, the results of several other sub-studies in this thesis provide strong evidence that the ASK is a unidimensional measure of disability consistent with clinicians', children's and other scales' definitions of disability. Furthermore, the literature on person-environment interaction suggests that disability is not restricted to the individual, but is an interaction of the person with the situation [Barris 1987, Cott et al. 1995, Haley et al. 1994a, Kielhofner 1993, Law 1991, Letts et al. 1994, Spencer 1993]. Thus, an alternative interpretation must be considered.

The second interpretation of these findings is that disability is affected by the situation in which it is measured. Although the ICIDH model implies that the impact of social and environmental factors is less significant at the disability level than at the handicap level, the ICIDH model does not exclude the influence of these factors on disability [WHO 1980]. If this interpretation is accepted, then the ASKp must continue to be considered within the disability domain and the results of the qualitative analysis may be considered to be a clarification of the role of contextual factors within disability. Thus, the role of situation in the ICIDH's definition of disability is not only supported by the quantitative and qualitative findings regarding the distinction between capability and performance in this thesis, but their role has been clarified by the identification of key themes and elements for one population. Additional research will be required to verify these themes and elements and explore those applicable to the definition of disability in other populations.

**How Capacity and Performance Relate to the ICIDH Model**

Given that both capability and performance lie within the domain of disability and that the qualitative interpretation of the difference clarifies the role of situation, the compatibility of this model of disability with respect to the literature must be addressed.
Information on the comparison of Nagi’s levels to the ICIDH levels is presented by Verbrugge and Jette in their paper “The Disablement Process” [Verbrugge and Jette 1994]. They clearly distinguish between situation-free “actions” and situation-dependent “activities”, and place both within the purview of the ICIDH’s disability level [Verbrugge and Jette 1994]. In their terms, the ASKc might be considered closer to situation-free, and belongs at the lower margin of the ICIDH’s disability level (perhaps within Nagi’s functional limitations level). The ASKp may be considered closer to situation-dependent, based on the qualitative results, and belongs at the upper margin of the ICIDH’s disability level (perhaps within Nagi’s disability level). Thus, the results of this thesis, integrated with the review of conceptual models, suggests that there may be a gradient effect, with greater situational or context effects occurring at the handicap level that are true for the impairment level. This interpretation is shown in Figure 10.

Figure 10 Integration of Disability Models

Kopec reworks the Nagi model and introduces a new perspective [Kopec 1995]. In his ‘Activity Space Model’ he too presents a level of situation-free activities that he terms elemental
abilities. His model is consistent with the trading-off concept identified in the qualitative analysis of the difference between capability and performance in this thesis, in that the actual performance of an activity is presented as a function of the impairment level combined with the external circumstances and psychological factors. He presents this visually as a difficulty curve, that he likens to the perceived exertion curve common to psycho-physical models. Although he also cites an important role for context in the expression of disablement [Kopec 1995], it is not clear whether Kopec proposes the activity space model as an explanation of the transition from impairment to disability (ICIDH) or disability to handicap. Thus, Verbrugge and Kopec's writings both appear to support the inclusion of a contextual component within disability.

Although Verbrugge and Jette's model of disability would suggest the ASKc should be situation-free, it was clear from the interviews that the children constructed a situation for the ASKc questions. Therefore, the ASKc was not truly situation-free. Some insight into why the ASKc was not situation-free is provided by Amartya Sen's writing on inequality and a review of the ASKc questionnaire (questionnaire is shown in Appendix VII). In a chapter titled "Functionings and Capability" Sen [1992] discusses capability and performance (functionings). Like Verbrugge, Sen discusses situation-free and situation-dependent aspects of disability. The term capability is associated with potential to function (or freedom to function) and is independent of the situation. Performance is associated with achieved functioning and is dependent upon the situation. Sen describes the relationship between performance and capability as follows: "achieved well-being [performance] itself depends on the capability to function" [page 41 Sen 1992]. Thus, capability is an essential precursor to physical performance and therefore, may only equal or exceed performance. This is consistent with the relationship between capability and performance observed in this thesis. So why did the children create a context for the freedom to function?

The explanation for the ASKc having a context, when the literature suggest it should be context-neutral, may be the result of the process of asking a capability question to a child. It
appears that the act of asking a child a capability question automatically invokes the mental construction of a situation. The wording of the ASKc further stimulated the creation of at least a constellation of symptoms (internal situation) by defining the time period of one week previous to the evaluation. Thus, while in theory the ASKc may have potentially been situation-free, this was not pragmatically possible. This is consistent with the gradient effect of context shown in Figure 10 in that the ASKc has a smaller context effect associated with it, but is not represented as being devoid of context (i.e., shaded grey rather than not being shaded at all). In summary, because the ASKp is influenced by factors encountered in the every day contexts of the children, it may be located closer to handicap than the ASKc. Because the ASKc attempted to minimize contextual factors, it may be located closer to impairment. This is depicted pictorially both in Figure 10 and Figure 11.

Implications for the Conceptual Basis of the ASK

The locations of the ASKp and the ASKc versions with respect to the ICIDH framework are important to the conceptual basis of the measure. McDowell and Newell [1996], among others [Haley et al. 1989, Nunnally and Bernstein 1994], have argued that a solid conceptual basis is critical in the development of any outcome measure. It is also pertinent to understand the location of the ASK with respect to alternative outcome measures. The review of literature described six paediatric outcome measures that had been reported to be indicators of childhood physical disability and were related in some way to the target population. These were the Childhood Health Assessment Questionnaire (CHAQ), Gross Motor Function Measure (GMFM), Health Utilities Index Mark III (HUI3), Juvenile Arthritis Functional Assessment Report (JAFAR), Pediatric Evaluation of Disability Inventory (PEDI), and Pediatric Functional Independence Measure (Wee FIM).

An analysis of the item content of these six measures with respect to the ICIDH definitions suggests that most contain a mix of impairment and disability items (e.g., the JAFAR), while some are more consistent with the definition of impairment than with disability.
(e.g., the GMFM). Those that do extend into disability are measuring capability exclusively, or the potential for function, rather than the expression of function in the usual context of children’s lives. Only the ASK-performance fully incorporates context effects. The location of each of these scales according to the ICIDH framework based on a review of their item content is shown in Figure 11.

**Figure 11 Conceptual Location of Paediatric Outcome Measures**

This diagram shows that the ASK versions occupy unique locations in the conceptual framework of consequences of disease. Thus, confirms the justification for developing the ASK. It also shows why the WeeFIM, HUB3, and GMFM are not acceptable measures of physical disability in this population because they focus on impairment. However, as previously stated, their inadequacy for measuring disability does not detract from their value in detecting impairment in their target populations. In fact, all three of these measures have undergone extensive testing on their target populations to document their validity.

Figure 11 above also indicates that the CHAQ, JAFAR and PEDI are conceptually the closest (of the six pre-existing measures) to a pure measure of childhood disability, although all
three contain a mix of impairment and disability items. This conceptual proximity of these three measures to the ASKc was not unexpected, since these measures were considered for modification and further development as an alternative to developing the ASK.

The PEDI was not modified because the age range and diagnostic group were significantly different from the intended population. The effects of growth and development between the two age ranges (six months to seven years for the PEDI and five to 15 years for the target population) were significant, and therefore modifying the PEDI was considered impractical. However, the ASK was developed with the intent of eventually creating a linkage between the ASK and the disability items within the PEDI. This would facilitate the use of the PEDI in younger children and the ASK in older children within the same study (e.g., for longitudinal studies). Therefore, the ASK was developed with a three year overlap in age (from five to seven years) with PEDI. This linkage may be achieved by item characteristic analysis of a combined data set in the future.

The CHAQ and JAFAR were also considered for modification, however, neither measure focused purely on disability. Furthermore, both were specific for children with arthritis. Moreover, both lacked a documented conceptual framework, and the review of their item content showed the CHAQ and the JAFAR had few child-specific items and few high level activities (e.g., running). Because of the “task” oriented approach of the JAFAR, it was considered to be slightly further from disability than was the CHAQ. Thus, although neither was appropriate for modification, the CHAQ was the closest to meeting the need for a disability measure among five to 15 year old children with musculoskeletal disorders and was included in the present research as a similar measure.

The results of this thesis have provided additional information on the measurement properties of the CHAQ and its relationship to the ASK. The data collected on the 200 children in this thesis demonstrated a rectangular distribution of CHAQ scores. This occurred in a sample of children reported to be skewed in terms of disability by the referring clinicians and
confirmed by the ASK and HUI3. Although all these reference criteria may be in error, it is unlikely. Furthermore, the distribution of disability among children is known to be skewed in the population [Hanvey et al. 1994a]. Thus, the rectangular distribution of CHAQ scores challenges its validity as a disability measure in this population. This inappropriate distribution was likely due in part to the aggregation methods. CHAQ scores include only the worst score in each domain, and penalty points are assigned for the use of assistive devices. Besides this challenge to the CHAQ’s validity, significant ceiling and floor effects were observed for the CHAQ. One final distinction between the CHAQ and the ASK is that the CHAQ is not formatted in a manner that is designed for child self-report. Previous pilot testing with the ASK underscored the importance of wording questions in the first person, and organizing them in coloured booklets with large type for child self-report. This likely has not been considered by the CHAQ’s developers, since it was developed from an adult measure and is usually completed by parent-report. This information further justifies the continued use of the ASK as opposed to the CHAQ among five to 15 year old children with musculoskeletal disorders.

Thus, the ASK is the most appropriate measure of disability for five to 15 year old children with musculoskeletal disorders. In addition to this distinction, it is the only measure that was developed directly from the perspectives of children. Accordingly, it contains the greatest proportion of child-specific questions (e.g., carry a back pack, keep up running) and has been designed to facilitate child self-report.

Implications for the Conceptualization and Measurement of Disability

The implications for these results to the broader field of disability evaluation in general can now be addressed. It has been argued that situation is an important aspect of disability measurement. Raphael [1996] reinforces this in his writing, and recommends that all quality of life evaluations include a qualitative component to capture the context when gathering quantitative data. Although the present research is not so far reaching in its interpretation (i.e., this research is about disability not quality of life) the results of this thesis do suggest
situational factors have a role in the measurement of disability. This role of context (or situation) in disability evaluation raises two questions. (1) Should situational variability influence the outcome of disability measures? (2) What are the implications of including a situational component in disability? This first question is not new. It is, in essence, asking where the dividing line is between disability and handicap [Badley 1987]. While the results here have made some suggestions of appropriate components of situation in measuring disability, the answer to this question must be addressed by further research. The current ICIDH revision process is currently attempting to address this question. However, there are several implications of including context into disability measures, that should be recognized.

The implications of including contextual factors in disability measurements requires detailed consideration. The fact that disability measured in a child's usual context is greater than when measured with an implied hypothetical context suggests that the usual situational contexts of childhood contain more negative factors (barriers) than positive factors (facilitators). This is consistent with the literature's focus on negative aspects of the environment to a much greater degree than on positive aspects of the environment (e.g., [Barris 1987, Kahana 1982, Kielhofner 1993, Law 1991, Letts et al. 1994, Swain et al. 1993]). Since the situation affects the outcome of disability measurement, it is important to consider situation as a relevant component in selecting an outcome measure. Thus, not only should measures be appropriate for the purpose and population, but also for the situation that we are trying to understand function in.

The identification of a trading-off process suggests a systematic integration of person and situation factors is responsible for performance. While the process is not entirely predictable, the situation factors may be equally important to the person factors for some aspects of disability. Thus, interventions directed at the situation may be highly successful in reducing disability. As a result, the acceptance of a person-situation interaction model of disability offers alternative treatment avenues.
The last issue with respect to the implications of the effect of situation on disability, is whether or not such an effect is appropriate. Because of the child-centred nature of the scale, not all items were necessarily applicable to all children. It was considered that this may have caused an effect apart from the situational effect the study was designed to explore. However, the implications of these not-applicable items were tested in this thesis and found to have no effect. Thus, the difference between capability and performance were ascribed to context effects.

Some questionnaires, such as the CHAQ have implied that context effects are inappropriate. This has been done by increasing children’s disability scores significantly (up to 66%) if assistive devices are used. However, even these apparently situation-free measures cannot achieve the freedom they espouse. For example, if a boy with severe arthritis dresses himself every day in a context in which assistive devices are prescribed, available, and used, he will score more disabled than a girl with the same degree of arthritis who achieves the same result but does so without any assistive devices. On the surface one might argue that the boy is more disabled. However, careful observation of physical behaviour patterns may show similar levels of disability. The girl who lacks assistive devices may score non-disabled on dressing items because she elects to put on a dress that has no buttons, zippers or ties. If assistive devices had been available she may have made the same trade-off as the boy did (i.e., use assistive devices and thus score greater disability). Essentially the girl has traded-off style for function in the absence of assistive devices, but she is really no less disabled than the boy. Thus, even measures that attempt to exclude context effects end up with variability in scores due to context (e.g., presence or absence of assistive devices).

This appropriateness of context in the conceptualization of disability was brought out in the review of literature where it was suggested that no human activity occurs in absolute isolation. Support for this statement came from a variety of sources [Barris 1987, Cott et al. 1995, Haley et al. 1994a, Kielhofner 1993]. Perhaps, most important among these was
Kielhofner’s assertion that disability is located at the interface between person and environment [Kielhofner 1993]. Thus, if there was no context there can be no disability evaluation.

In summary, if we accept that situation effects are conceptually part of disability and are clinically appropriate, then changes in the environment that result in increased ASK scores (defined as a reduction in disability) must be accepted as a true reduction in disability. Thus, if a child is fitted with a new prosthesis that allows him to participate in sports for the first time, the associated reduction in ASKp score must be accepted as a reduction in the child’s disability, not dismissed as artifact due to technological advances. It also follows that the existence of capability and performance versions of disability questionnaires may aid in the recognition of where situational factors are having the greatest influence on individual children with physical disabilities. For example, if a child is found to have moderate disability on the ASKp yet minimal disability on the ASKc, therapy directed at the pathology may be less beneficial than therapy directed at the environment. Therapy directed at both the pathology and the environment may be advised when both ASKp and ASKc scores indicate a similar degree of disability.

Now that the conceptual and clinical implications have been addressed, the methodologic importance of the present research and the interpretation of the ASK’s measurement properties may be addressed.

Integration of Multi-Method Approach at the Item Level

The methods and results chapters of this thesis addressed the measurement properties of the ASK and the impact of contextual factors on the expression of disability as two separate issues. To that end, the results documented the strong measurement properties of the ASK, and presented a theory on how the expression of disability is altered by context (the difference between capability and performance). However, the two aspects of the present research are intertwined. The ASK’s conceptual basis, item characteristics, clinimetric properties, psychometric properties, and context effects (predicted by the trading-off model) can now be
considered together and used to reassess the conceptual basis of the ASK’s items. This is demonstrated by revisiting the item content of the ASK. Through the use of two example item sets, the advantages of including children’s perspectives and using a multi-method research approach will become apparent.

Example 1

To begin with, the items related to personal care are considered. The items that represent the personal care domain were initially assigned to three domains (personal care, washing, and toileting). However, they have since been amalgamated into one (see Appendix III). These specific items were: turned on the tap, combed my hair, prepared and brushed my teeth, got on and off the toilet, used the toilet paper and flushed the toilet, got in and out of the tub or shower, and washed my whole body. Conceptually, these seven items were thought to represent basic activities of daily living. Since all items represent activities, and may be limited as a consequence of one or more physical impairments, they were all considered to represent the ICIDH construct of disability.

The Rasch model identified no misfit beyond the cut off for poor fit (0.52>0.1.7), but did identify a few minor areas of misfit (0.82>0.1.2) for the personal care items. The misfit was primarily due to redundancy, with most of the self-care activities (with the exception of tub transfers) being located at the easy end of the difficulty continuum that was explicated by the Rasch model. For example, the items getting on and off the toilet (ASK3 item four) and used the toilet paper and flushed the toilet (ASK3 item five) were located at logit values 0.43 and 0.46 respectively. These items have much in common (e.g., both relate to the same activity and have the same relative difficulty, although the biomechanical demands of each item are different). The Rasch model has provided an estimate of exactly how much they have in common, or how close their logit values are. This information was previously unknown. Based on the model, there were three items within one logit of item five, further suggesting
redundancy in this logit (difficulty) range. Thus, the Rasch model confirms and enriches the conceptual understanding of the ASK by identifying its item structure.

However, item characteristic analysis in isolation only provides part of the picture as to how the items relate to the overall conceptualization and measurement properties of the scale. The grouping of items into sub-domains is an important way to describe the components important to the face validity of the scale. The amount of error associated with individual items is also important (i.e., from item validity and responsiveness data). In this thesis Rasch analysis was applied in conjunction with clinimetric and psychometric data on items' validity, not-applicable rates and responsiveness (e.g., item five showed low responsiveness), to assess the ASK. The result was a comprehensive evaluation that permitted recommendations to enhance the conceptual foundation of the ASK without sacrifice of measurement properties.

The qualitative component of the present research also contributed information. Interviews with children identified a trading-off process that impacted on performance of self-care activities. For example, the availability of modifications (e.g., in terms of placement and height of fixtures in the bathroom) was reported to influence the performance of tooth brushing and toiletting. Thus, although the qualitative research did not alter the questionnaire's format it did enhance the understanding of the ASK's self-care domain.

In summary, the measurement properties and distinction between both capability and performance came together in the present research to evaluate a domain that was already conceptually sound, but that potentially contained some redundancy. While the redundancy is not of major concern, it does increase respondent burden and, if it is due to noisy items (e.g., brushing hair) it contains unnecessary amounts of error. The information from multiple methods provided a clear picture of the conceptual basis of the ASK and of which items were potentially problematic. There were other examples within the ASK in which the results of this thesis improved the conceptualization to a much greater degree than was possible prior to this analysis. The following section contains one such example.
Example 2

The domain of play (previously termed spare time activities) was one in which there was a minor degree of conceptual uncertainty as to which items truly belonged within the ASK. This domain included six items covering: quiet play activities (e.g., colouring with friends), heavy play activities (e.g., wrestling), independent sports (e.g., tennis), group sports (e.g., basket ball), gym class activities, and chores or hobbies (e.g., bed-making). Questions on children’s leisure time activities had not been included in previous childhood disability measures, however the previous measures had not considered children’s perspectives on disability. The decision to use the children’s perspectives was the key impetus to include a play or spare time activities domain in the ASK. However, given that these items had not previously been included in disability measures, the question was raised in a previous focus group as to whether play was a component of disability or handicap [Young 1994]. Because children expressed limitations in this domain in terms of activities, it was considered an aspect of disability in accordance with the ICIDH definition.

The Rasch results detected noise across many of the play activities. The qualitative research showed that the performance of play activities was strongly influenced by children’s affinity for activities, and by the presence of peers and the strength and direction of peers’ preferences. Conceptually, activities strongly affected by environmental factors may be more closely related to handicap than disability [Badley 1987, WHO 1980]. These combined findings of misfit and situational (or contextual) influence challenge the placement of play activities within the ICIDH’s domain of disability. The results from the Rasch model, the qualitative results, and the literature provide some insight into the conceptual location of these items and this domain. This conceptual information and information on clinimetric and psychometric properties of the ASK combine to inform us on the potential advantages and disadvantages of some items within the ASK’s spare time activities domain.
The Rasch model detected noise on both the infit and outfit statistics on some play items. This suggests that children both close to and far from these items' difficulty levels gave unexpected responses on these play items. This was most marked on the gym class activities question. In support of this finding, the qualitative interviews identified a great deal of variability in the degree of modification permitted within various children's gym classes. This was expressed in the trading-off model via the theme "modifications". Thus, in some cases, children with severe disability, who were permitted to use wheelchairs and given extra time in competitive gym class activities, may have participated to a greater degree than did children with lesser degrees of disability, for whom modifications were not available or permitted. Thus, the qualitative model suggests that extreme environmental influence may be one source of the misfit observed on the gym class activities item with respect to the Rasch model.

The first assumption behind Rasch analysis is unidimensionality. Since the principal construct was physical disability the items should fit the model if they belong to this construct. The finding of some misfit on the gym class activities item further questions the inclusion of this item as part of disability.

Other conceptual models were reviewed with respect to this challenge. These included the Nagi [1965] model, and modifications to the Nagi model suggested by Verbrugge et al. [1994] and Kopec [1995]. Nagi's model differed from the ICIDH model in that functional limitations were free of context while disability was a social construct, thus included context effects. Therefore, according to Nagi, play is a constituent of the highest level ('disability') rather than of functional limitations. However, since the trading-off process was observed across all activities, it seems that all items on the ASK performance belong in the highest level ('disability'). Since there was no higher level in which to place gym class activities, they must by default remain with the other items in Nagi's disability category. If we assume that Kopec's activity space model is a model of the transition from impairment to disability, then it supports the retention of play activities within the disability measure.
In summary, the retention of the play domain was weakly supported by the conceptual model and by modifications suggested by Verbrugge and Jette, and Kopec. The Rasch evidence of misfit was limited to some, but not all the play domain items. Perhaps those items with significant misfit are more strongly associated with handicap. Thus, the fit between the ASK play domain and the conceptual model of disability may be enhanced and threats to unidimensionality (posed by the inclusion of items potentially too closely related to handicap) reduced using the information provided by the multi-method approach used in this thesis to refine the item pool.

Overall, this review of two domains from the ASK has demonstrated by example how the two main objectives of the present research (assessment of measurement properties and exploration of contextual factors affecting performance) are intertwined. Thus, at the item level the value of information from the various sub-studies and methodologic approaches may be appreciated.

**Generalizability of Results and Limitations**

The results in this thesis have demonstrated the sound measurement properties of the ASK, identified the process that distinguishes between capability and performance, and demonstrated the value of an integrated multi-method approach to disability evaluation. The question that remains is how generalizable are these results?

**ASK Measurement Properties**

The results presented on the measurement properties of the ASK were conducted both on the total group of 200 children, and on smaller sub-samples. Where smaller sub-samples were used, descriptive data were presented to demonstrate the similarities between this group and the total sample to ensure generalizability to the total sample. The sample of 200 children covered the entire 5.0 to 15.99 year age range, and represented a variety of musculoskeletal conditions. Thus, it is anticipated that the results of the measurement properties reported in this study are
generalizable beyond the sample, to the general population of five to 15 year old children with musculoskeletal disorders.

There is, however, one caveat to this statement of generalizability of the ASK' s measurement properties. The responsiveness data were not intended to provide an estimate of effect sizes for a particular clinical group, but only to give an indication of the magnitude that might be detected in a group undergoing treatment of known effectiveness. Since the sub-sample of 34 children was heterogeneous, the actual effect size value cannot be used to calculate sample sizes for clinical trials. Rather, future studies are required to generate accurate effect size estimates for homogeneous populations of clinical interest.

**Conceptual Issues**

The results of this study also showed that capability consistently exceeded performance. This finding is consistent with previous reports from this research centre [Young et al. 1996] and others [Anderson et al. 1977]. The difference between capability and performance was slightly greater when performance scores are lower. As a consequence, the size of the difference is dependent to a small degree on the severity of disability in the population studies. Furthermore, the magnitude of the difference ranges from 8.6 to 35 points depending on whether binary or five point response options were used in the aggregation. However, it is not so much the magnitude of the difference as the significance and consistency regardless of method of aggregation. Because of the consistency of a significant difference across aggregation methods and across studies reported in the literature, it is expected that these results are generalizable beyond this sample of 200 children.

The qualitative component of the present research showed that the children went through a process of trading-off in determining their performance. There was consistency across activities, across children, and over time. However, caution is required in interpreting the results of qualitative research from a single sample based on single interviews. While the
Appendix III has recommended a reduction in items to increase feasibility. Thus, clinicians may now use a shorter 30 item ASK to document the disability and change in disability in a cost efficient and child-centred manner. Furthermore, they have the information necessary to help them decide which of the two versions is most appropriate for their own applications. Moreover, if both versions are applied they are expected to help clinicians identify areas in need of remediation. This remediation may take the form of therapy directed toward changing the person, or alteration of situational factors to eliminate barriers to function.

An important contribution was made by the qualitative component of the present research. The qualitative results uncovered some key aspects of the environment that may be addressed to reduce the expression of disability. Such information has not been previously identified by formal research methods.

The Rasch modeling approach used in this thesis represents the start of a data base that can be used as a reference group for further evaluations of disability in smaller groups. Specifically, these data represent an item anchor file [Linacre and Wright 1993] so that only those items close to a person’s ability level need to be answered in future applications, yet a summary score consistent with the completed 30-item administration can be later constructed statistically. This again offers efficiency, but will require further research to increase the
reference sample. In addition, the Rasch approach enables two distinct measures to be linked by analyzing a combined data set. The intent from the ASK's inception was that it would eventually be linked with the PEDI. Thus, the Rasch method used in this thesis will enable this to occur once a combined data set of ASK and PEDI data have been collected.

**Future Research into the Conceptualization of Disability**

This thesis has identified a trading-off process among children with physical disabilities. Although the main concept is supported by previous literature, replication of the results is recommended for future research. Furthermore, quantitative testing of other questionnaires reframed into capability and performance versions would be helpful in determining the generalizability of the capability-performance distinction in other questionnaires and other populations.

While the qualitative results of this thesis have been informative with regards to the conceptualization of disability, they must be considered exploratory. Thus, another qualitative study is necessary to verify the trading-off process reported in this thesis. It is expected that such research will enhance the understanding of the distinction between disability and handicap and the extent to which situations invade disability. Only then can the implications of the present research for the conceptualization of disability and the calculation of disability prevalence estimates be fully understood.
REFERENCES


Davis AM, Young NL, Williams JJ, Wright JG, Bombardier C and Bell R. Clinimetrics and Psychometrics: complimentary approaches to questionnaire development. Submitted to *Journal of Clinical Epidemiology* 1996.


Fries JF. The hierarchy of quality-of-life assessment, the Health Assessment Questionnaire (HAQ), and issues mandating development of a toxicity index. *Controlled Clinical Trials* 1991.


Rodwell MK. Person/environment construct: positivist versus naturalist, dilemma or opportunity for health social work research and practice?. [Review]. Social Science and Medicine 1990; 31(1): 27-34.


Wright BD and Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation* 1989; 70: 857-867.


Wright JG and Young NL. Comparison of different indices of responsiveness. Submitted to *Journal of Clinical Epidemiology* 1996.


APPENDICES

APPENDIX I  ALTERNATE DISABLEMENT MODELS

The ICIDH model [World Health Organization 1980], and the Nagi model [Nagi 1965] are both relatively recent developments. The ICIDH model remains popular in parts of Canada and throughout Europe, however, the Nagi model is most popular in the United States. In contrast to the ICIDH “medical model” (that classifies the consequences of disease [Duckworth 1984, Swain et al. 1993], the Nagi model was developed as a rehabilitation model, and the Verbrugge and Kopec papers add a socio-medical aspect to the model. Several authors have compared the ICIDH and Nagi models. Nagi’s response to these reviews has been that “The difference is not only in semantics nor simply that one framework is an extension of the other. There are also important substantive differences.” [pages 320-321 Nagi 1991]. However, there are common themes between these two primary models.

Nagi Model

In its initial 1965 description, the Nagi model had five levels: active pathology encompassed the body’s responses to infection or injury. In this framework impairments were defined as anatomical and/or physiological deviations, functional limitations were defined as restrictions affecting the performance of usual roles and daily activities, sickness and illness was defined as the behavioral responses to pathology, and disability was defined as the behavioral consequences of long-term impairment [Nagi 1965].

The model appears reasonable until examples are reviewed. Specifically, Nagi uses amputations as an example of an impairment “not involving active pathology” [page 102 Nagi 1965]. Furthermore, Nagi states that functional limitations must coincide with impairment. Thus, Nagi has restricted the linkages between categories on the basis of temporality (i.e., only
the current situation is considered). From a medical or rehabilitation perspective events that have gone before are of import. In addition, the distinction between sickness and illness and disability on the basis of duration appears artificial, although it may potentially be of clinical use. The sickness and illness category was dropped from Nagi’s 1991 publication [Nagi 1991]. It also appears that disability is almost exclusively behavioral in nature.

**Suggested Modifications of the Disablement Models**

The ICIDH and Nagi models have been discussed by many authors. However, few have offered a strong conceptual basis for their comments, or been constructive in suggesting modifications to the frameworks. Verbrugge and Jette [1994] and Kopec [1995] have recently published conceptual papers that provide food for thought with respect to the two primary models (ICIDH and Nagi models).

Verbrugge and Jette [1994] present a modified version of the Nagi model that has a much greater emphasis on psycho-social factors that feed into the disablement process. Their modification also presents a greater degree of inter-relatedness and feedback loops between levels of the Nagi model. Verbrugge and Jette conceptualize Nagi’s *functional limitations* and *disability* as being within the ICIDH category of *disability*. They equate *pathology* with diagnosis, *impairments* (indirectly) with signs and symptoms, *functional limitations* with actions (situation-free), and *disabilities* with activities (situation-dependent). Although some aspects of the revision appear artificial (e.g., the great emphasis on chronic versus acute diseases), there are some useful concepts. Particularly, the use of the term ‘action’ to define situation-independent events and ‘activity’ to define situation-dependent events clarifies the distinction between *functional limitations* (action events) and *disability* (activities) more adequately than Nagi’s terms of ‘concept’ and ‘relational concept’. Furthermore, the placement of Nagi’s concept of *disability* as lower on the hierarchy than handicap is intriguing, and leaves room for consideration of higher level restriction in discretionary aspects of daily life.
Kopec also presents an interesting conceptual addition to disability frameworks in what he terms the ‘Activity Space Model’ [Kopec 1995]. Kopec presents a level of situation-free activities that he terms *elemental abilities*. His model suggests that the actual performance of an activity is a function of the impairment level combined with the external circumstances and psychological factors. He presents this as a difficulty curve, that he likens to the perceived exertion curve common to psycho-physical models. Activities above the difficulty curve will be completed despite their cost because their value exceeds the associated cost, whereas activities below the curve will not be completed despite the capability to do so because of physical cost. This is a visual representation of a decision model that may explain the impact of situational factors, however, this is not well connected to the ICDH nor Nagi models. Thus, it is unclear whether Kopec proposes this as an explanation of the transition from *impairment* to *disability* (ICIDH) or *disability* to *handicap*. 
APPENDIX II  BACKGROUND ON RASCH MODELING

Introduction

Measures are used to quantify the degree of a trait or attribute such as disability. Physical
disability is a partially observable attribute, thus clinicians have been able to develop some
measures of disability. However, disability scales may benefit from the application of methods
used to develop measures to quantify less observable traits.

The conceptual knowledge essential to scale development for non-visible traits comes
primarily from psychology, where non-visible traits dominate research. Psychologists were
pioneers in the development of scaling methods that include Rasch modeling. Aspects of this
approach have also been adopted in education (perhaps introduced by psychologists interested
in intelligence testing), and later generalized to various forms of aptitude testing. Only recently
have the psychometric approaches been generalized to evaluations of health.

Historical Perspective

A historical review is one method for developing an understanding of the role of
psychometric methods in health evaluation. Thus, key events in the development of health
evaluation and psychometric methods are reviewed below.

Historical Review of Health Evaluation (via clinimetrics)

The need for numeric indicators of health arose for several reasons. Part of the impetus
came from an urgent need to find fast and cost-efficient methods to screen candidates for
military service in WWII. This need coincided with methodological advances in survey
sampling and data analysis that made numeric scales very advantageous. The methods needed
to be not only fast and efficient, but also valid for determining those of adequate physical health
to be trained for combat overseas [McDowell and Newell 1987]. The attribute to be measured
was physical function and thus was not entirely observable.
Little was known at the time about appropriate development and validation of measures of health or other intangible traits. Statistical strategies for determining item characteristics had not yet been developed, nor was there adequate computing power to perform the current day strategies had the methods for item characteristic analysis been available in the early 1940's. In that era, the logical source of information was from physicians. Because of their experience in the treatment of physical infirmity, physicians were expert diagnosticians, although not necessarily experts on health or quality of life. As a result, the first measures of health were developed based on clinical expertise (later termed ‘clinimetric’ strategies), in the absence of any substantial theories of measurement. These scales were validated by comparison to clinical judgment or clinical prediction of change.

Demographers used analogous methods to develop indicators of health, since advances in medicine had reached a level that indicators of mortality alone were no longer able to provide sufficient information. Thus, the knowledge learned from military screening was used in gathering census information that would provide information necessary for the description of the health status of developed countries in a similar way that mortality statistics are informative on the health status of third world countries.

The initial need for health status measures was to improve our ability to assess something other than quantity of life. Current medical care advances became unlikely to cause a change in quantity of life, thus, even more refined measures of quality of health components are important to evaluate therapeutic effectiveness [Streiner and Norman 1995]. The challenge of the 90's is a formidable one: to devise methods “to measure what was previously thought to be immeasurable, and assess in a reproducible and valid fashion those subjective states which cannot be converted into the position of a needle on a dial” [pages 1-2 Streiner and Norman 1995]. This may be beyond the scope of clinimetric methods, therefore investigation of alternative spheres of measurement knowledge such as psychometrics is warranted.
Historical Review of Psychometric Theory

Psychometrics was driven by a similar need to develop a method for evaluating large groups efficiently, but traveled along a more conceptual and theory-driven path. Psychometrics became firmly established and documented to the 1970’s (e.g., [Wright and Stone 1979, Nunnally 1978]). However, even after the advent of psychometrics an integration of clinimetric and psychometric strategies has not occurred to any great extent, either in a formal conceptual sense or a pragmatic sense. These two methods should not be looked upon as opposing strategies, but rather as distinct strategies, each of value and potentially very complementary in the process of scale development [Davis et al. Submitted].

Clinimetric strategies may be a very efficient method for generating items for a measure of health, but these strategies are unable to ascertain the characteristics of the items. Conversely, psychometric methods may be a very circuitous route to item generation, but are very efficient methods for determining the characteristics of items. For this reason, I have chosen psychometric methods to conduct additional testing on the Activities Scale for Kids (ASK), that was derived primarily along clinimetric guidelines.

Some of the first research on which psychometric theory was built was conducted by Weber in 1846. Weber’s work was considered to be in the domain of psycho-physics because he examined the perception of events in physics (e.g., light waves, weight, etc.). Weber was the first scientist to describe ‘just noticeable differences’, defined as the amount of change in a stimulus that was necessary for the change to be perceived subjectively. Weber’s Law stated that ‘just noticeable differences’ were proportional to the absolute level of the stimulus [McDowell and Newell 1987].

Weber’s research was the basis for Gustav Fechner’s research. In 1860 Fechner published his major work including Fechner’s law, that was derived from studies of stimulus-response characteristics of weight sensation, light intensity, etc. Fechner’s law proposed that the relationship between stimulus and response followed the natural logarithm model
[McDowell and Newell 1987], so that when the absolute level of a stimulus was high, extremely large increases were necessary for a change to be perceive, and when a stimulus was small and extremely small increase was perceptible.

Fullerton and Cattell (1892) were responsible for the link between psychophysics and psychometrics according to Nunnally [1994]. The Fullerton-Cattell law specified that ‘equal often notice differences are equal unless always or never noticed’ [Nunnally 1994]. This has meaning for preference responses used often in psychometrics, in that if armchairs are preferred over kitchen chairs 75% of the time, and kitchen chairs are preferred over bar stools 75% of the time, then the arm chair-kitchen chair distance is equivalent to the kitchen chair-bar stool difference. If one option (e.g., bar stools) were never preferred, then nothing could be said about the distance relationships between options.

It was a long time before Stevens’ power law was developed in the 1960’s as a refinement of Weber’s and Fechner’s work. Stevens added an exponent coefficient for the level of the stimulus, to account for different stimulus-response characteristics for different stimuli [McDowell and Newell 1987].

The nature of the stimulus-response relationship was also discussed with respect to the nature of the threshold. Thurston (1928) and Guttman (1944) proposed models of response that considered the threshold to be absolute [Edwards, 1957]. So although the amount of stimulus required to detect a change was related to the absolute value of the baseline stimulus logarithmically, the detection was an all or none phenomena. It was not until 1960 that Rasch proposed the detection model was not a step function, but an ogive (sigmoid shape). Thus, the probability of detection increased logarithmically with the location of the 50% probability point specified by Steven’s power law.

Classical test theory and Guttman scalogram analysis were both once the methods of choice in scale development. However, Guttman’s methods hinged on the principle that there was a threshold value before which all responses were positive and beyond which all answers were
negative. Empirical testing has shown response patterns variability near the transition point, thus response patterns are more probabilistic than deterministic (see Figure A1). Classical test theory fell out of favor with the advent of more sophisticated item response theory methods, primarily because it (like factor analysis and Guttman scalogram analysis) is dependent upon the composition of the sample. Thus, items selected as a good measure of a construct in one sample were not necessarily generalizable to other similar samples.

**Figure A1  Response Patterns & Probability Predictions**

<table>
<thead>
<tr>
<th>Case</th>
<th>Ideal Guttman Response Pattern</th>
<th>Typically Observed Response Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1 1 1 1 1 1 0</td>
<td>A 1 1 1 1 1</td>
</tr>
<tr>
<td>B</td>
<td>1 1 1 1 1 0 0</td>
<td>B 1 1 1 1</td>
</tr>
<tr>
<td>C</td>
<td>1 1 1 1 0 0 0</td>
<td>C 1 1 1</td>
</tr>
<tr>
<td>D</td>
<td>1 1 1 0 0 0 0</td>
<td>D 1 1</td>
</tr>
<tr>
<td>E</td>
<td>1 1 0 0 0 0 0</td>
<td>E 1</td>
</tr>
<tr>
<td>F</td>
<td>1 0 0 0 0 0 0</td>
<td>F 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>Guttman Probability Predictions</th>
<th>Rasch Probability Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1 1 1 1 1 1 0</td>
<td>A .99 .9 .8 .7 .6</td>
</tr>
<tr>
<td>B</td>
<td>1 1 1 1 1 0 0</td>
<td>B .95 .85 .75 .65 .35 .3</td>
</tr>
<tr>
<td>C</td>
<td>1 1 1 1 0 0 0</td>
<td>C .9 .8 .7 .35 .25</td>
</tr>
<tr>
<td>D</td>
<td>1 1 1 0 0 0 0</td>
<td>D .85 .75 .3 .2 .1</td>
</tr>
<tr>
<td>E</td>
<td>1 1 0 0 0 0 0</td>
<td>E .7 .3 .2 .1 .01</td>
</tr>
<tr>
<td>F</td>
<td>1 0 0 0 0 0 0</td>
<td>F .3 .2 .1 .01</td>
</tr>
</tbody>
</table>

The identification of the natural logarithm model was important to the work of George Rasch, who was also busily publishing in the 1960's on his item response theory model. The Rasch model determined the probability of success on an item, given an underlying absolute level of an attribute. Like stimulus-response relationships, this relationship was also a natural logarithm. Stevens' earlier work on classifying numeric scales as nominal, ordinal, interval and ratio, and provided rules to govern the mathematical processes applicable to scales of different numeric natures (e.g., ordinal data can only be appropriately be analyzed by rank
analysis methods) [page 11 Nunnally and Bernstein 1994]. These rules were also a building block of Rasch analysis. One of the benefits of Rasch modeling is its ability to transform ordinal data to interval level data, thus facilitate its analysis with more powerful parametric statistics.

Subsequent work on validation of the power law has contributed to our understanding of scaling. Because subjective judgments have traditionally been given numerical values, work was eventually done to understand the values given. This work subsequently lead to the use of "magnitude estimation" to validate ordinal scales [McDowell and Newell 1987].

Selection of a Specific Psychometric Approach

Types of statistics may be broken into descriptive and inferential. Descriptive statistics concern quantitative statements about a group of observations that are not necessarily generalizable. Inferential statistics concern generalizations from a set of observations to the larger population, usually in the form of probability statements. Descriptive methods include primarily correlational analysis and factor analysis that are useful in understanding the relationships between items. Cluster analysis may be considered a variant of factor analysis, and is also concerned with the relationships between items. Profile analysis and discriminant analysis are similar to classical factor analysis, however, are more interested in scaling persons than items. Rasch analysis and other item characteristic analysis methods fit within the domain of inferential statistics because of their generalizability.

Factor Analysis

The most familiar method of psychometric analysis is factor analysis. Factor analysis looks at the correlations between items within a scale to determine which items are most closely related. Closely related items are grouped together, and are said to 'load' on a common factor. There may be multiple factors within a scale, and one item may load upon more than one factor. However, it is sometimes difficulty to discern the nature of a factor. This may be due in part to the non-specific nature of the reference point in factor analysis; the origin or
perspective from which the factors are viewed is arbitrary. Rotations and subsequent analyses are often performed to confirm the initial analysis, on the premise that if the same factor loadings are generated from two different vantage points, then they likely represent appropriate factor loadings. Furthermore, factor analysis cannot determine the magnitude of the attribute measured, only the strength with which it is associated with other items in the pool. Moreover, factor analysis is limited in its generalizability beyond the initial sample of subjects. Thus, the factor loadings from one sample may bear little relationship to the factor loadings from another sample.

Classical psychometric theory has relied predominantly on factor analysis and other forms of descriptive rather than inferential statistics. Descriptive statistics use closed-form solutions rather than open-form (e.g., iterative) solutions, the latter requiring complex computing due to the iterative nature of open form calculations. Thus, for many years the practical limitations of computers constrained psychometricians to descriptive statistics. Inferential methods have advanced in the recent past due to the computing power improvements that gave birth to maximum-likelihood estimation methods. Since the purpose of scale development is to measure more than one sample, and generalization is important, inferential statistics have a definite advantage. Modern psychometric methods, that utilize inferential statistics to generate open form solutions, are based on item response theories. The most common and simplest of the item response theory models is the one-parameter Rasch model. There are also two- and three-parameter models.

**Item Response Theory & Item Characteristic Analysis**

The greatest benefit offered by inferential statistics such as the item response theory models, is that they can simultaneously consider the scaling of subjects and of items. Thus, not only are the results of a sample of 25 subjects' responses to 10 items generalizable to other subjects, but they also provide information about how these subjects would perform on five items on which they were not tested (given that the items have known item characteristics from
previous Rasch modeling application). This ability to simultaneously model persons and items permits the results to be free of sample bias. The results of the modeling are generalizable, and contain information not only about how strongly items are related, but also on where they are located along the continuum of interest (location is expressed as a logit value between negative four and positive four), and how sharply they discriminate (discriminative ability is described by the slope in the two- and three-parameter models).

**History of Item Characteristic Modeling**

This history of item characteristic modeling was drawn from a variety of sources, primarily the Measurement Transactions published by MESA Press at the University of Chicago, with some additional information on its relationship to other methods coming from Nunnally [Nunnally and Bernstein 1994].

Item characteristic models are generated by statistical methods that analyze response patterns to questions. Initially these were developed for binary responses, but methods of analyzing ordinal multiple choice items have also been developed. As mentioned previously, item characteristic analysis has its basis in psychological test development, building on the concept of just-noticeable-difference and stimulus-response relationships described by Weber, Fechner, and Stevens.

The first text on Rasch analysis was published by Georg Rasch in 1960. Georg Rasch (1901-1980) was a Danish mathematician who began this approach in order to analyze responses to reading tests in 1953. However, most of the history of the Rasch model is tied to the University of Chicago, where Rasch was a visiting professor in 1959. Benjamin Wright PhD, who was to become a keystone in the development of the method, was a professor teaching statistics in the Education department at the University of Chicago. Although unfamiliar with Rasch’s work, Wright was mildly intrigued and, more importantly, felt morally obligated to attend the lecture series. His interest was spawned by frustration with the inconsistencies (variance) he experienced with using factor analysis methods on repeated
educational tests. Thus, there were at least three crucial linkages with educational scaling research that contributed to a method that is now being used in health assessment and psychology. However, there was also an interesting linkage to the maths as Wright had done his undergraduate training in physics, and Mike Linacre (who was to follow) had a degree in mathematics.

During the visiting professorship Rasch and Wright forged a relationship that seemingly disappeared in the spring of 1960, but was rekindled by Wright's desire to travel to Denmark to get Rasch's perspective on some semantic differential data in 1964. Gradually, Wright began to see that correlational and factor analysis methods were unwieldy and unnecessary confusing. In contrast, Rasch's model was clean and clear. Wright's interest was maintained and stimulated in Chicago by a series of students and fellows who were involved in theoretical testing, generation of simulated data sets, and writing of computer programs to share the knowledge. One of these students was John Michael Linacre, who obtained a PhD under Wright's tutelage, and wrote the BIGSTEPS program that superseded BICAL, CREDIT, MScale and MSTEPS and is the current day standard.

As Wright and his students at the University of Chicago continued to work diligently on variations of the Georg Rasch's ideas and generate computer programs to perform the analysis, a similar group under Gerhart Fischer were working along similar lines in Europe. For reasons unknown to Wright's group item response theory and Rasch modeling was more readily adopted in Europe and remains more common place there than in North America. However, it has begun to take hold in education primarily in North America and there are some advantages to the North American methods (e.g., much larger data sets can be analyzed). The Rasch model has been applied to several health measures in the last 10 years, including the: FIM, Tufts Assessment of Motor Performance, the PEDI and later the SF36.
Methods

The basis of the Rasch (one-parameter) model and two- and three-parameter logistic models is that the response patterns of persons on a series of items are compared. The response patterns of items answered by a series of persons are also compared simultaneously. This results in a model that is based on one dimension (that being measured by the test) that reports the ability level of subjects and difficulty level of items. The ability levels of subjects are test-free in that they are not dependent upon the particular subset of items administered. Similarly, the item difficulty levels are sample-free in that they are not dependent upon the group of subjects tested. It is this simultaneous modeling of subjects and items via maximum-likelihood estimation that is unique to item response theory (IRT) methods when compared to traditional descriptive psychometric methods such as factor analysis.

All IRT methods are based on three assumptions: (1) that the data are unidimensional, (2) that there is local independence among items after accounting for the dominant dimension (e.g., disability), and (3) that there is an independence from the sample. These assumptions are tested in the modeling process [Hambleton 1991, Molenaar 1995]. Lack of unidimensionality is obvious by high infit and outfit item statistics for items not belonging to the dominant dimension.

Item Characteristic curves are monotonic that plot the probability of a correct answer on the y-axis and ability on the x-axis (technically the x-axis is the difference between item difficulty ($b_i$) and person ability ($\theta$), where the average person ability is zero). Thus, the probability of a correct answer increases with an increase in the underlying trait. The conceptual basis comes from latent trait theory, and originally item characteristic curves (ICC’s) were proposed to be related to the normal distribution [Streiner and Norman 1989], now believed to be logistic (close to a cumulative normal function). Thus, the curves have an ogive shape (S shape). Curves are not permitted to overlap as this runs contrary to construct validity. The steepness of the ogive is an indicator of its discriminative ability, because a steep
curve indicates a large change in the probability for a small change in the difference between the item difficulty level and the person’s ability level. Thus, ICC’s with slopes greater than one discriminate well. An item’s difficulty position is specified by the logit value at the point where the ICC crossed the 50% probability (see Figure A2 below). Similarly a person’s ability level is expressed as a logit or log-odds.

The person and item parameters are expressed in logits or log-odds ratio units because the model is logistic. As Wright states: it is easier for us to use logits which then appear linear and behave in a linear fashion than to use untransformed values that demand us to think of logistic relationships (our eyes pick out straight lines very well, logistic curves very poorly) [Wright 1996]. A logit is the log of the probability of success divided by the probability of failure. Logits are not universal in that the logits from one scale cannot be directly equated with those from another, unless data exist where some items from each scale were administered to the same sample of subjects. This is how the Rasch model equates tests, confirms language transformations of tests, facilitates item banking, and enables tests to be tailored to the level of the subjects without loss of information about the items not administered.

All three primary models are summarized below, based on the descriptions of Hambleton 1991]. Several modifications are discussed in the Rasch transactions.

**one-parameter model:** the only factor affecting the ICC’s is the item difficulty denoted $b_i$, the ICC’s all have the same slope $a$ where $a = 1$, thus $a$ can effectively be eliminated from equation 1, the ICC’s all have the same origin $c = zero$.

Thus, the probability of a given response frequency by a group of subjects is determined by their ability level $\theta$ and the item difficulty $b_i$.

$$Pi (\theta) = \frac{e^{(\theta - b_i)}}{1 + e}$$

**one-parameter formula:**
two-parameter model: the ICC's are still affected by item difficulty ($b_i$), but

the ICC's each have their own slope thus $a$ is given a subscript $i$,
the ICC's continue to have the same origin ($c = 0$). 

Thus, the probability of a given response frequency by a group of subjects is determined by their ability level $\theta$, the item difficulty ($b_i$), and the slope ($a_i$)

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

three-parameter model: the ICC's are still affected by item difficulty denoted $b_i$, but

the ICC's each have their own slope ($a_i$), and
the ICC's each have their own origin ($c_i = 0$).

Thus, the probability of a given response frequency by a group of subjects is determined by their ability level $\theta$, the item difficulty ($b_i$), the same slope ($a_i$) and origin ($c_i$).

$$P_i(\theta) = c_i + (1 - c_i) x \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

Streiner and Norman stated that "Models have been developed to handle items with more than two responses, but they are computationally quite difficult, and few computer programs exist to handle these cases" [page 139 Streiner and Norman 1989], and [page 183 Streiner and Norman 1995]. In fact, practical applications do exist to analyze multiple choice responses with a one-parameter model, but both the two- and three-parameter models to be problematic for several reasons. For example, they both require much larger sample sizes than the sample sizes of 50 to 150 recommended by Wright and Linacre [Wright and Stone 1979, Linacre 1994, Wright 1996] or even the 200 recommended by Streiner & Norman [1995], and Numnally[1994]. The estimated sample is in the range of 1000 subjects [Streiner and Norman 1995], and thus is definitely not feasible in medical research. Even with such a large sample convergence often does not occur.
Furthermore, some of the specifics of health measurement scales negate the benefits of the more complicated two- and three-parameter models. For example, the use of multiple response options, rather than binary, effectively increases the slope of the item response function and causes all slopes to approach one. When this occurs (as is the case in the ASK), slight differences in slope are less essential to model, and potential benefits of two-parameter models become obsolete. Moreover, in educational settings, the effect of guessing makes it highly improbable that any subject would score zero. However, in clinical medicine, it is conceivable that a subject’s physical ability to complete daily activities could be zero (e.g., vegetative state). Thus, the need to permit intercepts other than zero, the main benefit of the three-parameter model, is not relevant for health care measures. Therefore, from both a pragmatic and theoretical perspective the most appropriate model for health measure development is the one-parameter model, also termed the Rasch model.

As mentioned in the methods chapter, the Rasch model satisfies the aim of freedom from sample bias [Molenaar 1995]. Freedom from sample and item biases is important because it permits generalization across items for any subject. This means that one can explain responses to items, predict the probability of a given response on items from the estimates of ability level of the subject (based on the subjects’ responses to other items) and the estimate of difficulty level of the items (based on other subjects’ responses) without ever observing their disability on all possible items [Molenaar 1995]. Furthermore, the Rasch model permits generalization across persons in a similar way. Because not all items must be answered, tests can be shortened (or targeted as is done with mastery tests) without loss of information. Since it is generalizable across subjects, a representative sample for test development is favorable but not essential [Molenaar 1995].

Hypothetical item characteristic curves generated by the Rasch model for five items are shown in Figure A2. The curves are normal ogive curves that show the probability success based on the physical ability level of the subject. In this diagram, walking and keeping up with peers discriminate between patients with low and moderate levels of ability with more
confidence than climbing stairs. Since the intent was to maximize discrimination at all levels of physical disability item selection sought to retain a set of items with steep ogive curves spaced equidistantly across the physical disability continuum.

The position of an item with respect to other items is determined by the location of the middle response option at the point on the curve where the probability is equal to 0.5. As shown in Figure A2, climbing stairs has a difficulty level of 0.2. The Bigsteps® software is a Rasch modeling program. It is able to identify and plot the position and shape of the ogive curves for each item independently (groups=0 command denotes response options are distinct for each item) or plot them assuming similar response patterns for all items (default option treats response categories similarly across all items). The latter plots are generated only after the assumption has been verified with groups=0 testing.

Figure A2 Sample Item Response Curves
Summary

It is clear from the historical review presented that factor analysis and descriptive statistics have been superseded in their role in test development by the advent of super-computers and maximum likelihood estimation and inferential statistics. Nunnally states that “any serious student of psychometrics will need to know the fundamentals of IRT even though classical procedures will suffice in most applications” [page 393 Nunnally and Bernstein 1994]. IRT modeling is by far the optimal method of the day with which to test the psychometric properties of a scale, and will complement the clinimetric work done to data in completing the development of the ASK. Given the instability and convergence problems inherent in the two- and three-parameter IRT models, the one-parameter Rasch model was the logical choice.

It appears that the Rasch model has much to offer in the development and application of new measures. The benefits of Rasch application to the ASK were as follows:

1 to identify the structure of the ASK (and in the process confirm unidimensionality); specifically the location of items according to level of difficulty, so that a range of items that are equally spaced across the trait may be selected,

2 to identify misfitting items to be improved or removed,

3 to test the stability of item parameters between the ASKc and ASKp versions,

4 to prepare for the intended linking with the Pediatric Evaluation of Disability Inventory scores in the future, and

5 to prepare for the possibility of diagnosis-specific tailored tests and item banking in the more distant future.
APPENDIX III  ITEM HISTORY

Preliminary Item Reduction - Methods

In order to conduct the comparisons described in this thesis, data on both the ASKp and ASKc were needed from a large sample of children. Since each version of the ASK contained 73 multiple-choice items, it was advisable to minimize the demands placed on these children. Therefore, a preliminary item reduction strategy was used. This process was an attempt to strike a balance between the risk of poor compliance due to high respondent burden and the risk of deleting items too early. In order to prevent the deletion of important items prematurely, the intent was to reduce the ASK to no fewer than 50 items.

Data from previous research were used to make this interim reduction in the number of items. This reduction was performed by consensus of four experts who reviewed the data previously collected on the 73 item ASKp and ASKc from 28 children. These data included item frequencies, importance ratings, and item position data from an exploratory Rasch model. Thus, although there was the potential to lose valuable items through this process, the risk of losing a key item was considered remote due to the large number of items to be considered and the wealth of information of various types available from previous research.

Preliminary Item Reduction - Results

Data previously collected using the 73 item ASK from 28 children [Young 1994]. provided the necessary information for this process. Two main criteria were used to identify items for deletion. First, those items that detected no disability on the ASKc, ASKp, or relative need response options on the ASKp, or had a not-applicable rate above 70% were identified. This strategy resulted in the identification of 15 items. Second, those items that were conceptually related to another similar item (e.g., going up a curb, and going down a curb) were identified. This strategy resulted in the identification of an additional 14 items.
These 29 items were reviewed with respect to their importance values (from previous item
generation interviews) and difficulty levels (from a preliminary Rasch model). Based on this
review, 23 of the 29 items identified were deleted (17 dropped completely, six integrated into
revised items), and six were retained (two were challenging items, three represented other
items that had been dropped, one underwent a major revision). This left 50 items, to which
three new items were added. These were added for the following reasons: one item was
double-barreled (ASK73 item 15, put on socks and shoes) and required subdivision (in ASK53
items 12 and 13), and two were included to reduce the previously apparent [Young 1994]
ceiling effect by the addition of more challenging items (ASK53 items 38 and 40: ran and
played outside with friends, and participated in gym class, respectively). The data to support
this reduction are shown in Table A1 of Appendix III. A description of each of the 73 items is
presented in Table A6 in Appendix III.

Review of the ASKp data also indicated that there were 55 items that, although children
reported performing less frequently than seven days per week, were meeting their needs. This
review found 26 items (24 of which were also reported among the 55 just mentioned) that,
although children reported performing them seven days per week, were not meeting their
needs. This suggest that the response options used on the ASK73 (zero to seven days per
week) were problematic because sometimes activities required more than daily performance for
the child to be considered non-disabled while for other activities less than daily performance
was not necessarily indicative of disability. Thus, the final 53 items were given judgmental
response options (none of the time, a little of the time, sometimes, most of the time, and all of
the time) in the version used to collect data for the present research.

The data collected using the ASK33 were later subjected to Rasch analysis. The diagnostic
information from this analysis and other analyses of validity suggested a further refinement of
the ASK was possible. Based on this information, a 30 item version was proposed. The
progression from the 73 item ASK to the final proposed ASK30 is shown in Table A6.
<table>
<thead>
<tr>
<th>ASK73 item</th>
<th>Percent with Limitation</th>
<th>Needs Met</th>
<th>Value (importance)</th>
<th>Difficulty (ASKc</th>
<th>ASKp</th>
<th>Decision</th>
<th>ASK53 item</th>
<th>Disability Sub-Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>84</td>
<td>78</td>
<td>0</td>
<td>77</td>
<td>3.6</td>
<td>-0.06</td>
<td>0.17</td>
<td>retained</td>
</tr>
<tr>
<td>2</td>
<td>76</td>
<td>65</td>
<td>7</td>
<td>71</td>
<td>2.0</td>
<td>0.32</td>
<td>0.62</td>
<td>retained</td>
</tr>
<tr>
<td>3</td>
<td>66</td>
<td>66</td>
<td>57</td>
<td>67</td>
<td>0.1</td>
<td>0.79</td>
<td>1.05</td>
<td>dropped*</td>
</tr>
<tr>
<td>4</td>
<td>88</td>
<td>85</td>
<td>3</td>
<td>84</td>
<td>2.0</td>
<td>0.00</td>
<td>-0.22</td>
<td>retained</td>
</tr>
<tr>
<td>5</td>
<td>84</td>
<td>65</td>
<td>17</td>
<td>81</td>
<td>3.1</td>
<td>0.00</td>
<td>0.36</td>
<td>integrated</td>
</tr>
<tr>
<td>6</td>
<td>86</td>
<td>77</td>
<td>35</td>
<td>94</td>
<td>3.1</td>
<td>-0.16</td>
<td>0.52</td>
<td>retained</td>
</tr>
<tr>
<td>7</td>
<td>69</td>
<td>66</td>
<td>3</td>
<td>76</td>
<td>3.6</td>
<td>0.42</td>
<td>0.62</td>
<td>retained</td>
</tr>
<tr>
<td>8</td>
<td>84</td>
<td>66</td>
<td>3</td>
<td>84</td>
<td>3.6</td>
<td>-0.12</td>
<td>0.59</td>
<td>dropped*</td>
</tr>
<tr>
<td>9</td>
<td>92</td>
<td>85</td>
<td>3</td>
<td>92</td>
<td>4.5</td>
<td>-0.25</td>
<td>-0.47</td>
<td>retained</td>
</tr>
<tr>
<td>10</td>
<td>88</td>
<td>85</td>
<td>3</td>
<td>84</td>
<td>4.5</td>
<td>-0.25</td>
<td>-0.16</td>
<td>dropped*</td>
</tr>
<tr>
<td>11</td>
<td>96</td>
<td>92</td>
<td>7</td>
<td>92</td>
<td>4.5</td>
<td>-0.65</td>
<td>-0.60</td>
<td>retained</td>
</tr>
<tr>
<td>12</td>
<td>73</td>
<td>75</td>
<td>0</td>
<td>73</td>
<td>7.6</td>
<td>0.32</td>
<td>0.29</td>
<td>retained</td>
</tr>
<tr>
<td>13</td>
<td>88</td>
<td>82</td>
<td>0</td>
<td>81</td>
<td>7.6</td>
<td>-0.33</td>
<td>-0.28</td>
<td>retained</td>
</tr>
<tr>
<td>14</td>
<td>80</td>
<td>78</td>
<td>0</td>
<td>58</td>
<td>7.6</td>
<td>-0.25</td>
<td>-0.22</td>
<td>retained</td>
</tr>
<tr>
<td>15</td>
<td>76</td>
<td>85</td>
<td>3</td>
<td>72</td>
<td>7.6</td>
<td>0.37</td>
<td>0.04</td>
<td>revised</td>
</tr>
<tr>
<td>16</td>
<td>84</td>
<td>88</td>
<td>7</td>
<td>83</td>
<td>7.6</td>
<td>0.16</td>
<td>-0.34</td>
<td>revised</td>
</tr>
<tr>
<td>17</td>
<td>88</td>
<td>80</td>
<td>10</td>
<td>88</td>
<td>3.1</td>
<td>0.00</td>
<td>0.05</td>
<td>retained</td>
</tr>
<tr>
<td>18</td>
<td>88</td>
<td>47</td>
<td>32</td>
<td>83</td>
<td>3.1</td>
<td>-0.12</td>
<td>0.18</td>
<td>retained</td>
</tr>
<tr>
<td>19</td>
<td>80</td>
<td>46</td>
<td>53</td>
<td>91</td>
<td>1.2</td>
<td>0.22</td>
<td>0.67</td>
<td>integrated</td>
</tr>
<tr>
<td>20</td>
<td>76</td>
<td>0</td>
<td>75</td>
<td>71</td>
<td>1.2</td>
<td>0.37</td>
<td>2.57</td>
<td>revised</td>
</tr>
<tr>
<td>21</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>1.8</td>
<td>-1.69</td>
<td>-2.29</td>
<td>dropped</td>
</tr>
<tr>
<td>22</td>
<td>80</td>
<td>90</td>
<td>21</td>
<td>95</td>
<td>1.8</td>
<td>0.32</td>
<td>-0.56</td>
<td>integrated</td>
</tr>
<tr>
<td>23</td>
<td>87</td>
<td>80</td>
<td>46</td>
<td>86</td>
<td>1.8</td>
<td>-0.11</td>
<td>0.15</td>
<td>revised</td>
</tr>
<tr>
<td>24</td>
<td>83</td>
<td>81</td>
<td>3</td>
<td>88</td>
<td>1.8</td>
<td>-0.23</td>
<td>-0.21</td>
<td>revised</td>
</tr>
<tr>
<td>25</td>
<td>84</td>
<td>75</td>
<td>0</td>
<td>77</td>
<td>1.8</td>
<td>-0.18</td>
<td>-0.01</td>
<td>retained</td>
</tr>
<tr>
<td>26</td>
<td>88</td>
<td>78</td>
<td>0</td>
<td>77</td>
<td>1.8</td>
<td>-0.11</td>
<td>-0.01</td>
<td>retained</td>
</tr>
<tr>
<td>27</td>
<td>69</td>
<td>38</td>
<td>35</td>
<td>44</td>
<td>0.1</td>
<td>0.66</td>
<td>1.15</td>
<td>retained</td>
</tr>
<tr>
<td>28</td>
<td>100</td>
<td>70</td>
<td>3</td>
<td>92</td>
<td>0.2</td>
<td>-1.69</td>
<td>0.31</td>
<td>dropped</td>
</tr>
<tr>
<td>29</td>
<td>80</td>
<td>84</td>
<td>7</td>
<td>88</td>
<td>5.3</td>
<td>0.09</td>
<td>-0.25</td>
<td>retained</td>
</tr>
<tr>
<td>30</td>
<td>88</td>
<td>92</td>
<td>3</td>
<td>92</td>
<td>5.3</td>
<td>-0.33</td>
<td>-0.83</td>
<td>retained</td>
</tr>
<tr>
<td>31</td>
<td>92</td>
<td>80</td>
<td>7</td>
<td>96</td>
<td>5.3</td>
<td>-0.41</td>
<td>-0.45</td>
<td>retained</td>
</tr>
<tr>
<td>32</td>
<td>84</td>
<td>75</td>
<td>0</td>
<td>88</td>
<td>5.3</td>
<td>-0.25</td>
<td>-0.17</td>
<td>retained</td>
</tr>
<tr>
<td>33</td>
<td>80</td>
<td>80</td>
<td>10</td>
<td>96</td>
<td>5.3</td>
<td>2.59</td>
<td>-0.43</td>
<td>retained</td>
</tr>
<tr>
<td>34</td>
<td>84</td>
<td>64</td>
<td>39</td>
<td>100</td>
<td>6.4</td>
<td>0.00</td>
<td>0.53</td>
<td>retained</td>
</tr>
<tr>
<td>35</td>
<td>76</td>
<td>75</td>
<td>14</td>
<td>96</td>
<td>6.4</td>
<td>0.00</td>
<td>0.05</td>
<td>revised</td>
</tr>
<tr>
<td>36</td>
<td>76</td>
<td>66</td>
<td>35</td>
<td>94</td>
<td>6.4</td>
<td>0.17</td>
<td>0.54</td>
<td>integrated</td>
</tr>
<tr>
<td>37</td>
<td>80</td>
<td>84</td>
<td>32</td>
<td>100</td>
<td>6.4</td>
<td>0.00</td>
<td>0.16</td>
<td>revised</td>
</tr>
<tr>
<td>38</td>
<td>65</td>
<td>72</td>
<td>35</td>
<td>100</td>
<td>6.4</td>
<td>0.75</td>
<td>0.38</td>
<td>retained</td>
</tr>
</tbody>
</table>

continued...
<table>
<thead>
<tr>
<th>4SK73 item</th>
<th>Percent with Limitation</th>
<th>Value (importance)</th>
<th>Difficulty</th>
<th>Decision</th>
<th>4SK53 item</th>
<th>Disability Sub-Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>96 82 0 100</td>
<td>1.9 -0.52 -0.34</td>
<td>dropped</td>
<td>31</td>
<td>Carrying</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>84 80 7 88</td>
<td>1.9 -0.12 0.08</td>
<td>retained</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>84 63 21 90</td>
<td>1.9 0.11 0.28</td>
<td>revised</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>88 64 10 100</td>
<td>1.9 -0.12 -0.02</td>
<td>retained</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>80 77 3 88</td>
<td>3.4 0.37 0.12</td>
<td>revised</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>80 76 7 91</td>
<td>3.4 0.37 0.18</td>
<td>integrated*</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>80 72 10 91</td>
<td>3.4 0.37 0.13</td>
<td>revised</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>80 64 10 90</td>
<td>3.4 0.42 0.31</td>
<td>integrated*</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>76 68 21 95</td>
<td>3.4 0.52 0.39</td>
<td>retained</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>76 84 10 87</td>
<td>3.4 0.56 0.09</td>
<td>dropped*</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>87 62 71 88</td>
<td>3.4 -0.46 0.28</td>
<td>dropped</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>91 75 71 100</td>
<td>3.4 -0.79 -0.19</td>
<td>dropped</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>100 85 3 92</td>
<td>9.9 -1.61* -0.71</td>
<td>dropped</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>96 61 7 96</td>
<td>9.9 -1.16 0.24</td>
<td>retained</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>70 52 10 81</td>
<td>7.8 0.87 0.84</td>
<td>retained</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>66 30 7 82</td>
<td>7.8 0.92 0.96</td>
<td>retained</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>78 55 67 89</td>
<td>1.7 0.47 0.66</td>
<td>retained</td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>92 92 3 92</td>
<td>3.1 -0.11 -0.55</td>
<td>retained</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>87 66 78 67</td>
<td>3.1 0.10 -0.09</td>
<td>dropped</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>92 84 10 87</td>
<td>0.3 -0.79 -0.60</td>
<td>retained</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>96 74 3 92</td>
<td>0.3 -1.16 -0.46</td>
<td>retained</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>96 96 0 100</td>
<td>3.1 -1.16 -1.78</td>
<td>retained</td>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>100 92 0 100</td>
<td>1.1 -1.61* -0.96</td>
<td>dropped*</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>96 92 3 96</td>
<td>0.8 -1.16 -1.35</td>
<td>dropped*</td>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>92 96 3 100</td>
<td>1.1 -0.36 -1.68</td>
<td>dropped*</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>96 81 3 96</td>
<td>1.1 -1.16 -0.34</td>
<td>retained</td>
<td>56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>83 84 7 96</td>
<td>3.1 0.16 -0.14</td>
<td>retained</td>
<td>57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>84 70 3 79</td>
<td>3.7 -0.04 0.04</td>
<td>retained</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>80 74 3 76</td>
<td>3.7 0.28 0.17</td>
<td>dropped*</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>66 66 46 92</td>
<td>3.7 1.06 0.71</td>
<td>revised</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>92 78 0 88</td>
<td>0.5 -0.60 -0.22</td>
<td>retained</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>84 82 0 85</td>
<td>1.1 0.15 -0.06</td>
<td>revised</td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>96 96 0 96</td>
<td>1.1 -0.47 -0.96</td>
<td>dropped*</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>92 71 0 95</td>
<td>1.1 -0.19 -0.06</td>
<td>retained</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>92 92 0 92</td>
<td>1.1 -0.11 -0.73</td>
<td>dropped*</td>
<td>65</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Excessive misfit to Rasch model
* Conceptual overlap with retained item
The table on the previous pages shows the information used to refine the ASK73 to the ASK53. The ASK53 was used to collect the data for the current research. During the present research more information on item behaviour was gained through item characteristic analysis and validity testing. It was anticipated that as a consequence of this information, some potential refinements to the ASK may be identified. These were primarily expected to come from the diagnostic analysis done by the Rasch model. However, the Rasch analysis was only one of several strategies that could be integrated into the selection of a smaller item set for future research. The final decision to remove or modify items also integrated information on item validity, responsiveness, and from a review of the items' conceptual fit to the ICIDH model. The item responsiveness data are presented in Table A2.

Table A2 Individual Item Responsiveness (based on five point response options)

<table>
<thead>
<tr>
<th>Item</th>
<th>ASKp &amp; ASKc Improvement Combined</th>
<th>ASKp Improvement</th>
<th>ASKc Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Effect Size</td>
</tr>
<tr>
<td>1</td>
<td>0.56</td>
<td>1.35</td>
<td>0.42</td>
</tr>
<tr>
<td>2</td>
<td>0.24</td>
<td>0.70</td>
<td>0.34</td>
</tr>
<tr>
<td>3</td>
<td>0.51</td>
<td>1.17</td>
<td>0.43</td>
</tr>
<tr>
<td>4</td>
<td>0.82</td>
<td>1.52</td>
<td>0.54</td>
</tr>
<tr>
<td>5</td>
<td>0.57</td>
<td>1.57</td>
<td>0.36</td>
</tr>
<tr>
<td>6</td>
<td>1.45</td>
<td>1.72</td>
<td>0.84</td>
</tr>
<tr>
<td>7</td>
<td>1.03</td>
<td>1.53</td>
<td>0.67</td>
</tr>
<tr>
<td>8</td>
<td>1.21</td>
<td>1.48</td>
<td>0.82</td>
</tr>
<tr>
<td>9</td>
<td>0.55</td>
<td>0.88</td>
<td>0.63</td>
</tr>
<tr>
<td>10</td>
<td>1.09</td>
<td>1.55</td>
<td>0.70</td>
</tr>
<tr>
<td>11</td>
<td>0.58</td>
<td>1.05</td>
<td>0.55</td>
</tr>
<tr>
<td>12</td>
<td>1.00</td>
<td>1.38</td>
<td>0.73</td>
</tr>
<tr>
<td>13</td>
<td>0.85</td>
<td>1.39</td>
<td>0.61</td>
</tr>
<tr>
<td>14</td>
<td>1.27</td>
<td>1.70</td>
<td>0.75</td>
</tr>
<tr>
<td>15</td>
<td>1.18</td>
<td>1.83</td>
<td>0.65</td>
</tr>
<tr>
<td>16</td>
<td>0.90</td>
<td>1.28</td>
<td>0.70</td>
</tr>
<tr>
<td>17</td>
<td>0.28</td>
<td>0.81</td>
<td>0.35</td>
</tr>
<tr>
<td>18</td>
<td>0.13</td>
<td>0.64</td>
<td>0.21</td>
</tr>
<tr>
<td>19</td>
<td>1.06</td>
<td>1.37</td>
<td>0.78</td>
</tr>
<tr>
<td>20</td>
<td>0.72</td>
<td>1.31</td>
<td>0.55</td>
</tr>
<tr>
<td>21</td>
<td>0.36</td>
<td>1.61</td>
<td>0.23</td>
</tr>
</tbody>
</table>

continued...
<table>
<thead>
<tr>
<th>Item</th>
<th>ASKp &amp; ASKc Improvement Combined</th>
<th>ASKp Improvement</th>
<th>ASKc Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Effect Size</td>
</tr>
<tr>
<td>22</td>
<td>1.07</td>
<td>1.93</td>
<td>0.56</td>
</tr>
<tr>
<td>23</td>
<td>0.81</td>
<td>1.50</td>
<td>0.54</td>
</tr>
<tr>
<td>24</td>
<td>1.45</td>
<td>1.72</td>
<td>0.84</td>
</tr>
<tr>
<td>25</td>
<td>1.11</td>
<td>1.44</td>
<td>0.77</td>
</tr>
<tr>
<td>26</td>
<td>1.39</td>
<td>1.67</td>
<td>0.83</td>
</tr>
<tr>
<td>27</td>
<td>1.04</td>
<td>1.93</td>
<td>0.54</td>
</tr>
<tr>
<td>28</td>
<td>1.49</td>
<td>1.71</td>
<td>0.87</td>
</tr>
<tr>
<td>29</td>
<td>1.08</td>
<td>1.90</td>
<td>0.57</td>
</tr>
<tr>
<td>30</td>
<td>0.80</td>
<td>1.24</td>
<td>0.65</td>
</tr>
<tr>
<td>31</td>
<td>0.83</td>
<td>1.47</td>
<td>0.56</td>
</tr>
<tr>
<td>32</td>
<td>1.28</td>
<td>1.57</td>
<td>0.81</td>
</tr>
<tr>
<td>33</td>
<td>0.39</td>
<td>1.40</td>
<td>0.28</td>
</tr>
<tr>
<td>34</td>
<td>0.98</td>
<td>1.61</td>
<td>0.61</td>
</tr>
<tr>
<td>35</td>
<td>1.19</td>
<td>1.73</td>
<td>0.69</td>
</tr>
<tr>
<td>36</td>
<td>1.56</td>
<td>1.77</td>
<td>0.88</td>
</tr>
<tr>
<td>37</td>
<td>0.41</td>
<td>0.95</td>
<td>0.43</td>
</tr>
<tr>
<td>38</td>
<td>1.41</td>
<td>1.59</td>
<td>0.89</td>
</tr>
<tr>
<td>39</td>
<td>1.83</td>
<td>1.72</td>
<td>1.06</td>
</tr>
<tr>
<td>40</td>
<td>1.55</td>
<td>2.17</td>
<td>0.71</td>
</tr>
<tr>
<td>41</td>
<td>1.57</td>
<td>1.84</td>
<td>0.85</td>
</tr>
<tr>
<td>42</td>
<td>1.63</td>
<td>1.91</td>
<td>0.85</td>
</tr>
<tr>
<td>43</td>
<td>0.78</td>
<td>1.52</td>
<td>0.52</td>
</tr>
<tr>
<td>44</td>
<td>0.86</td>
<td>1.62</td>
<td>0.53</td>
</tr>
<tr>
<td>45</td>
<td>1.05</td>
<td>1.53</td>
<td>0.69</td>
</tr>
<tr>
<td>46</td>
<td>0.81</td>
<td>1.46</td>
<td>0.55</td>
</tr>
<tr>
<td>47</td>
<td>0.87</td>
<td>1.64</td>
<td>0.53</td>
</tr>
<tr>
<td>48</td>
<td>1.16</td>
<td>1.63</td>
<td>0.71</td>
</tr>
<tr>
<td>49</td>
<td>1.36</td>
<td>2.03</td>
<td>0.67</td>
</tr>
<tr>
<td>50</td>
<td>1.47</td>
<td>1.39</td>
<td>1.06</td>
</tr>
<tr>
<td>51</td>
<td>0.85</td>
<td>1.36</td>
<td>0.63</td>
</tr>
<tr>
<td>52</td>
<td>0.69</td>
<td>1.45</td>
<td>0.48</td>
</tr>
<tr>
<td>53</td>
<td>1.22</td>
<td>1.46</td>
<td>0.83</td>
</tr>
</tbody>
</table>

The final item selection to optimize the ASK’s measurement properties was achieved by summarizing the item diagnostics with data previously collected on the items according to the criteria listed in Table A3. These criteria are listed in approximate order of importance, where
the first four criteria were particularly important, but no single criteria was sufficient to cause removal of an item on its own. Subjective judgment of goodness of fit to the ICIDH conceptual model of disability, and domain of the item were also considered in the final decision. Thus, the final item selection attempted to provide the best possible combination items that: (1) maximize the spread of items along the difficulty continuum, (2) maximize discrimination between individuals across the five to 15 year age range, and (3) have strong evidence for validity and responsiveness.

Table A3  Item Selection Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Rating</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise</td>
<td>x</td>
<td>excessive variability - fit statistic &gt; 1.7</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>elevated variability - fit statistic &gt; 1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[Wright and Linacre 1994]</td>
</tr>
<tr>
<td>Redundancy</td>
<td>x</td>
<td>too little variability - fit statistic &lt; 0.5</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>low variability - fit statistic &lt; 0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[Wright and Linacre 1994]</td>
</tr>
<tr>
<td>Location</td>
<td>numeric value</td>
<td>logit value -4 to +4</td>
</tr>
<tr>
<td>Clustering of items</td>
<td>numeric value</td>
<td>number of items within ±0.1 logit</td>
</tr>
<tr>
<td>Relationship to overall score</td>
<td>x</td>
<td>point biserial &lt; 0.25</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>*</td>
<td>Effect Size ≥ 1.00</td>
</tr>
<tr>
<td></td>
<td>x</td>
<td>Effect Size &lt; 0.25</td>
</tr>
<tr>
<td>Criterion Validity</td>
<td>x</td>
<td>ICC &lt; 0.40</td>
</tr>
</tbody>
</table>

* additional information on infit and outfit statistics can be found in the review of literature section on Item Characteristic Analysis

Table A4 shows a summary of the analysis of item characteristics resulting from the application of a grouped Rasch model with three response options. The letter following the item number indicates it's sub-domain (A through P, see Appendix III Table A5 for descriptions of items and sub-domains). Items listed in bold were questioned during the preliminary item reduction, but retained, and therefore required critical review at this point. Misfit was defined as a fit statistic less than 0.5 (redundant) or greater than 1.7 (noisy), and as
questionable fit if less than 0.8 (redundant) or greater than 1.2 (noisy) on either version [Wright and Linacre 1994]. Location is the Rasch model logit value, with a potential range of negative four to positive four. Two location values are given, the first is for the ASKp and the second, in brackets, is for the ASKc model. Clustering of items was defined as the number of items within ±0.1 logits. Relation to overall score was assessed by the point biserial from the Rasch output, and reported if less than 0.25. Responsiveness was considered inadequate if the effect size was less than 0.5. (Note that this criteria was initially set at 0.25, however, all items met this criteria, therefore it was raised.) Criterion validity was inadequate if Kappa was below 0.4. The final decision was a subjective integration of this material, taking into consideration the domain to which the item belonged to prevent elimination of domains from the scale, and conceptual fit (a subjective judgment as to the appropriateness of the item with respect to the World Health Organization's model of disability [WHO 1980]).

Table A4  Item Final Refinement

<table>
<thead>
<tr>
<th>Item</th>
<th>Misfit Noise</th>
<th>Misfit Redundant</th>
<th>Relation to total</th>
<th>Item Location</th>
<th>Clustering of items</th>
<th>Responsiveness</th>
<th>Criterion Validity</th>
<th>Final Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 F</td>
<td>x (x)</td>
<td>x (x)</td>
<td>2.35 (2.61)</td>
<td>0</td>
<td>unable to calculate</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 F</td>
<td>? (?)</td>
<td></td>
<td>1.25 (1.01)</td>
<td>2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 I</td>
<td></td>
<td></td>
<td>1.21 (1.20)</td>
<td>2</td>
<td>retained</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 E</td>
<td>x (x)</td>
<td>x (x)</td>
<td>1.16 (0.17)</td>
<td>2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 M</td>
<td>x (x)</td>
<td>x (x)</td>
<td>0.89 (1.25)</td>
<td>1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 L</td>
<td>? (x)</td>
<td></td>
<td>0.79 (1.64)</td>
<td>3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 D</td>
<td></td>
<td></td>
<td>0.77 (0.78)</td>
<td>2</td>
<td>unable to calculate</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43 M</td>
<td>?</td>
<td></td>
<td>0.74 (0.46)</td>
<td>2</td>
<td>retained</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52 O</td>
<td>?</td>
<td></td>
<td>0.63 (0.79)</td>
<td>3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 A</td>
<td>(x)</td>
<td></td>
<td>0.61 (1.29)</td>
<td>3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 A</td>
<td>(x)</td>
<td></td>
<td>0.61 (0.87)</td>
<td>3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46 N</td>
<td>?</td>
<td></td>
<td>0.58 (0.54)</td>
<td>4</td>
<td>retained</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 B</td>
<td></td>
<td></td>
<td>0.48 (0.49)</td>
<td>3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 I</td>
<td></td>
<td></td>
<td>0.46 (0.35)</td>
<td>2</td>
<td>retained</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Item numbers in bold indicate those items questioned during the preliminary item reduction.

x denotes an item of poor quality according to this criteria
(x) denotes poor quality on ASKc
? denotes an item of questionable quality according to this criteria
(?) denotes questionable quality on ASKc
* denotes an item of exceptional quality according to this criteria
<table>
<thead>
<tr>
<th>Item</th>
<th>Misfit Noise</th>
<th>Misfit Redundant</th>
<th>Relation to total</th>
<th>Item Location</th>
<th>Clustering of items</th>
<th>Responsiveness</th>
<th>Criterion Validity</th>
<th>Final Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fit statistic</td>
<td>fit statistic</td>
<td>point biserial</td>
<td>logit value</td>
<td># of items within</td>
<td>Effect Size</td>
<td>Kappa</td>
<td></td>
</tr>
<tr>
<td>fitting</td>
<td>$x &gt; 1.7$</td>
<td>$x &lt; 0.5$</td>
<td>$&lt; 0.25$</td>
<td>ASKp (ASKc)</td>
<td>$\pm 0.1$ logit</td>
<td>$x &lt; 0.5$</td>
<td>$x &gt; 0.4$</td>
<td>retained or $x = deleted$</td>
</tr>
<tr>
<td>4B</td>
<td>?</td>
<td></td>
<td>0.43 (0.21)</td>
<td>2</td>
<td>$x$</td>
<td></td>
<td></td>
<td>retained</td>
</tr>
<tr>
<td>51O</td>
<td>?</td>
<td></td>
<td>0.28 (0.40)</td>
<td>2</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33J</td>
<td>?</td>
<td></td>
<td>0.27 (0.54)</td>
<td>4</td>
<td>N/A $&lt;$ 25%</td>
<td></td>
<td></td>
<td>retained</td>
</tr>
<tr>
<td>44M</td>
<td>?</td>
<td>(?)</td>
<td>0.19 (0.78)</td>
<td>6</td>
<td>$x$</td>
<td></td>
<td></td>
<td>retained$^1$</td>
</tr>
<tr>
<td>7C</td>
<td>?</td>
<td></td>
<td>0.17 (0.21)</td>
<td>6</td>
<td>$x$</td>
<td></td>
<td></td>
<td>retained</td>
</tr>
<tr>
<td>27I</td>
<td>?</td>
<td></td>
<td>0.17 (-0.17)</td>
<td>6</td>
<td>N/A $&lt;$ 25%</td>
<td></td>
<td></td>
<td>retained$^1$</td>
</tr>
<tr>
<td>34K</td>
<td>?</td>
<td></td>
<td>0.10 (0.19)</td>
<td>6</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10D</td>
<td>?</td>
<td></td>
<td>0.09 (-0.11)</td>
<td>7</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11D</td>
<td>?</td>
<td></td>
<td>0.05 (-0.07)</td>
<td>7</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2A</td>
<td>?</td>
<td>(?)</td>
<td>0.00 (0.06)</td>
<td>9</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25I</td>
<td>?</td>
<td></td>
<td>-0.01 (-0.12)</td>
<td>9</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47N</td>
<td>?</td>
<td></td>
<td>-0.04 (-0.37)</td>
<td>6</td>
<td>$x$</td>
<td></td>
<td></td>
<td>retained$^1$</td>
</tr>
<tr>
<td>26I</td>
<td>?</td>
<td></td>
<td>-0.05 (-0.10)</td>
<td>7</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48N</td>
<td>?</td>
<td></td>
<td>-0.06 (-0.76)</td>
<td>8</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12D</td>
<td>?</td>
<td></td>
<td>-0.08 (-0.15)</td>
<td>7</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30I</td>
<td>(x)</td>
<td></td>
<td>-0.11 (-0.17)</td>
<td>6</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28I</td>
<td>?</td>
<td></td>
<td>-0.16 (-0.32)</td>
<td>4</td>
<td>N/A $&lt;$ 25%</td>
<td></td>
<td></td>
<td>retained$^1$</td>
</tr>
<tr>
<td>31J</td>
<td>?</td>
<td></td>
<td>-0.23 (-0.19)</td>
<td>2</td>
<td>$x$</td>
<td></td>
<td></td>
<td>retained$^1$</td>
</tr>
<tr>
<td>35K</td>
<td>?</td>
<td></td>
<td>-0.29 (-0.88)</td>
<td>3</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21H</td>
<td>(?)</td>
<td></td>
<td>-0.34 (-0.48)</td>
<td>2</td>
<td>N/A $&lt;$ 25%</td>
<td></td>
<td></td>
<td>retained$^1$</td>
</tr>
<tr>
<td>20G</td>
<td>?</td>
<td></td>
<td>-0.39 (0.01)</td>
<td>3</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13D</td>
<td>?</td>
<td></td>
<td>-0.47 (-0.52)</td>
<td>8</td>
<td>$x$</td>
<td></td>
<td></td>
<td>retained$^1$</td>
</tr>
<tr>
<td>15E</td>
<td>?</td>
<td></td>
<td>-0.51 (-0.81)</td>
<td>8</td>
<td>N/A $&lt;$ 25%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22I</td>
<td>?</td>
<td></td>
<td>-0.52 (-0.94)</td>
<td>8</td>
<td>$x$</td>
<td></td>
<td></td>
<td>retained$^1$</td>
</tr>
<tr>
<td>32J</td>
<td>?</td>
<td></td>
<td>-0.53 (-0.70)</td>
<td>8</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6C</td>
<td>?</td>
<td></td>
<td>-0.55 (-0.85)</td>
<td>8</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50O</td>
<td>?</td>
<td></td>
<td>-0.55 (-0.58)</td>
<td>8</td>
<td>N/A $&lt;$ 25%</td>
<td></td>
<td></td>
<td>retained$^1$</td>
</tr>
<tr>
<td>14E</td>
<td>?</td>
<td></td>
<td>-0.55 (-0.63)</td>
<td>8</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8D</td>
<td>?</td>
<td></td>
<td>-0.57 (-0.00)</td>
<td>10</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53P</td>
<td>?</td>
<td></td>
<td>-0.60 (-0.71)</td>
<td>10</td>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Item numbers in bold indicate those items questioned during the preliminary item reduction.

$x$ denotes an item of poor quality according to this criteria 

(x) denotes poor quality on ASKc

? denotes an item of questionable quality according to this criteria

(?) denotes questionable quality on ASKc

* denotes an item of exceptional quality according to this criteria

1 Initially identified for deletion, but conceptual importance outweighed evidence for deletion.
### Table A4 Final Item Refinement continued

<table>
<thead>
<tr>
<th>Item</th>
<th>Misfit Noise</th>
<th>Misfit Redundant</th>
<th>Relation to total</th>
<th>Item Location</th>
<th>Clustering of items</th>
<th>Responsiveness</th>
<th>Criterion Validity</th>
<th>Final Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fit statistic</td>
<td>fit statistic</td>
<td>point biserial</td>
<td>logit value</td>
<td># of items within</td>
<td>Effect Size</td>
<td>Kappa</td>
<td>retained or deleted</td>
</tr>
<tr>
<td></td>
<td>$x &gt; 1.7$</td>
<td>$x &lt; 0.5$</td>
<td>$&lt; 0.25$</td>
<td>$\text{ASK_p (ASK_c)}$</td>
<td>$\pm 0.1 \text{logit}$</td>
<td>$x &lt; 0.5$</td>
<td>$x &lt; 0.4$</td>
<td></td>
</tr>
<tr>
<td>38 L</td>
<td>?</td>
<td>(?)</td>
<td>-0.66 (-0.95)</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>36 K</td>
<td>?</td>
<td>(?)</td>
<td>-0.67 (-1.10)</td>
<td>5</td>
<td>N/A mm &gt; 25%</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 L</td>
<td>x (x)</td>
<td>-0.69 (-0.94)</td>
<td>5</td>
<td>N/A mm &gt; 25%</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42 L</td>
<td>?</td>
<td>(?)</td>
<td>-0.74 (-0.79)</td>
<td>4</td>
<td>N/A mm &gt; 25%</td>
<td>retained</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29 I</td>
<td>?</td>
<td>(?)</td>
<td>-0.80 (-0.85)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>retained</td>
</tr>
<tr>
<td>19 G</td>
<td>?</td>
<td>(?)</td>
<td>-0.93 (-0.47)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>retained</td>
</tr>
<tr>
<td>39 L</td>
<td>?</td>
<td>(?)</td>
<td>-0.93 (-0.97)</td>
<td>1</td>
<td>*</td>
<td>x</td>
<td>retained</td>
<td></td>
</tr>
<tr>
<td>49 N</td>
<td>?</td>
<td>(?)</td>
<td>-1.06 (-0.43)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>retained</td>
</tr>
<tr>
<td>41 L</td>
<td>x</td>
<td>(?)</td>
<td>-1.29 (-1.20)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

Item numbers in bold indicate those items questioned during the preliminary item reduction.  
$x$ denotes an item of poor quality according to this criteria.  
$(x)$ denotes poor quality on ASKc.  
? denotes an item of questionable quality according to this criteria.  
(?) denotes questionable quality on ASKc.  
* denotes an item of exceptional quality according to this criteria.

Initially 31 items were considered for deletion, two of which were recommended primarily on the basis of many items at the same difficulty level (ASK33 items five and 15). In addition to this concern, one had a large proportion of not-applicable responses on the ASKp33 (item 15) and the other had low responsiveness (item five). After considering the domain to which the items slated for deletion belonged, the other items being deleted, and their conceptual contribution to the ASK 24 items were recommended for deletion. These items are indicated in the table above. Two exceptions were made in this process. First, ASK33 item 21 (took care of special medical needs) was retained because it was considered by children to be very important to their independence and represented a conceptually unique activity, not represented by any other item. This had a questionable fit statistics that may in part be due to a low number of responses on these questions (ASK33 item 21 had a high not-applicable rate). Second, one new item (ASK30 item 11) was added because recent interviews with the children indicated the need for a printing item. Preliminary testing of this item suggests it is also highly responsive to clinically important change.
The result was a 30 item version of the ASK (ASK30) with nine domains. The domains (and number of items per domain) were as follows: personal care (three), dressing (four), eating and drinking (one), locomotion (seven) play (two), stairs (one), standing skills (five), transfers (five), and miscellaneous (two). The descriptions of these items are shown in Table A6 of this appendix. This final item set was analyzed to determine whether deletion of items had resulted in any substantial change in the scales measurement properties. Since the primary rationale for the selection of items was to improve fit to the Rasch model, this was assessed first. The result was a slight improvement in fit. The final model is shown in order of increasing item difficulty (easy items at the top) in Table A5.

Spearman's correlations with the CHAQ were also calculated using the shorter ASK30, and found to be 0.80 for the ASK30 performance and 0.81 for the ASK30 capability. These are comparable to the correlations between the CHAQ and ASK53. The responsiveness data were recalculated using the 30 items. The effect sizes were found to be 1.56 and 1.66 for the ASK30 performance and 1.10 and 1.75 for the ASK30 capability among those worsening and those improving, respectively. These statistics represent a slight (average of <0.1 standardized units) but non-significant decline in the responsiveness over the values reported for the ASK53. Although reliability is adversely affected by a reduction in item number, this effect is greatest when fewer than 20 items are retained [Nunnally and Bernstein 1994]. Therefore, a significant reduction in reliability is not expected based on theory, but remains to be determined empirically.

In summary, the selection of 30 items resulted in a more efficient measure (43% reduction in length) with no loss of validity or responsiveness, and a slight improvement in predictability of the response pattern. The history of each item from the initial ASK73, through the ASK53 to the ASK30 is shown in Table A6 on the next page.
<table>
<thead>
<tr>
<th>Item Description</th>
<th>Logit</th>
<th>Std. Error</th>
<th>Infit</th>
<th>Outfit</th>
</tr>
</thead>
<tbody>
<tr>
<td>23 got around inside my home</td>
<td>1.44</td>
<td>0.16</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>9 put my shirt on</td>
<td>0.97</td>
<td>0.14</td>
<td>0.97</td>
<td>0.87</td>
</tr>
<tr>
<td>43 got in &amp; out of a chair</td>
<td>0.93</td>
<td>0.14</td>
<td>0.72</td>
<td>0.48</td>
</tr>
<tr>
<td>3 prepared and brushed my teeth</td>
<td>0.80</td>
<td>0.14</td>
<td>0.89</td>
<td>0.92</td>
</tr>
<tr>
<td>46 got in &amp; out of bed</td>
<td>0.76</td>
<td>0.14</td>
<td>0.74</td>
<td>0.53</td>
</tr>
<tr>
<td>24 walked in crowds</td>
<td>0.62</td>
<td>0.15</td>
<td>1.07</td>
<td>0.98</td>
</tr>
<tr>
<td>4 got on &amp; off the toilet</td>
<td>0.60</td>
<td>0.14</td>
<td>0.91</td>
<td>0.64</td>
</tr>
<tr>
<td>44 sat on the floor</td>
<td>0.34</td>
<td>0.14</td>
<td>1.21</td>
<td>1.45</td>
</tr>
<tr>
<td>27 walked up a hill</td>
<td>0.33</td>
<td>0.15</td>
<td>1.07</td>
<td>0.89</td>
</tr>
<tr>
<td>7 washed my whole body</td>
<td>0.31</td>
<td>0.13</td>
<td>0.98</td>
<td>1.09</td>
</tr>
<tr>
<td>10 put my pants on</td>
<td>0.23</td>
<td>0.13</td>
<td>0.72</td>
<td>0.62</td>
</tr>
<tr>
<td>11 fastened my clothes (buttons &amp; zippers)</td>
<td>0.17</td>
<td>0.14</td>
<td>1.05</td>
<td>1.01</td>
</tr>
<tr>
<td>47 got down onto the floor &amp; up again</td>
<td>0.08</td>
<td>0.13</td>
<td>0.99</td>
<td>0.82</td>
</tr>
<tr>
<td>26 got around outside</td>
<td>0.08</td>
<td>0.13</td>
<td>1.26</td>
<td>1.12</td>
</tr>
<tr>
<td>X printed or wrote</td>
<td>-0.03</td>
<td>0.80</td>
<td>0.39</td>
<td>0.28</td>
</tr>
<tr>
<td>28 walked on rough or slippery surfaces</td>
<td>-0.04</td>
<td>0.15</td>
<td>1.12</td>
<td>1.08</td>
</tr>
<tr>
<td>31 carried a drink or food to the table</td>
<td>-0.13</td>
<td>0.13</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td>35 climbed up &amp; down a flight of stairs</td>
<td>-0.19</td>
<td>0.14</td>
<td>0.92</td>
<td>0.76</td>
</tr>
<tr>
<td>21 took care of my medical needs</td>
<td>-0.24</td>
<td>0.14</td>
<td>1.32</td>
<td>1.34</td>
</tr>
<tr>
<td>13 put my shoes on and did them up</td>
<td>-0.39</td>
<td>0.13</td>
<td>1.06</td>
<td>1.05</td>
</tr>
<tr>
<td>22 walk without any support</td>
<td>-0.44</td>
<td>0.12</td>
<td>1.25</td>
<td>1.17</td>
</tr>
<tr>
<td>32 carried things in 2 hands</td>
<td>-0.46</td>
<td>0.13</td>
<td>0.98</td>
<td>0.92</td>
</tr>
<tr>
<td>14 got in &amp; out of an automobile</td>
<td>-0.48</td>
<td>0.13</td>
<td>0.73</td>
<td>0.62</td>
</tr>
<tr>
<td>50 stretched to reach a high shelf</td>
<td>-0.50</td>
<td>0.15</td>
<td>0.77</td>
<td>1.04</td>
</tr>
<tr>
<td>53 got through heavy doors</td>
<td>-0.53</td>
<td>0.12</td>
<td>0.71</td>
<td>0.73</td>
</tr>
<tr>
<td>42 did my usual job or chores</td>
<td>-0.71</td>
<td>0.16</td>
<td>0.78</td>
<td>0.84</td>
</tr>
<tr>
<td>29 ran/rolled outdoors &amp; kept up with friends</td>
<td>-0.75</td>
<td>0.13</td>
<td>1.46</td>
<td>1.94</td>
</tr>
<tr>
<td>19 made a snack</td>
<td>-0.88</td>
<td>0.13</td>
<td>0.74</td>
<td>0.70</td>
</tr>
<tr>
<td>39 played sports I enjoy by myself</td>
<td>-0.88</td>
<td>0.13</td>
<td>1.56</td>
<td>1.99</td>
</tr>
<tr>
<td>48 stood still for 10 minutes</td>
<td>-1.02</td>
<td>0.13</td>
<td>1.15</td>
<td>1.32</td>
</tr>
</tbody>
</table>

Infit statistics > 1.8 indicate noise among the responses of children close to the ability level of the item. Outfit statistics > 1.8 indicate noise among the responses of children far from the ability level of the item. Infit or outfit statistics < 0.5 indicate redundancy between this and other items.

1 This is a new item added in the final revision. Its position in the Rasch model is estimated based on eight responses. Its inclusion in the model does not alter the location of the other 29 items.
<table>
<thead>
<tr>
<th>Sub-Domain</th>
<th>ASK73</th>
<th>ASK53</th>
<th>ASK30</th>
<th>Sub-Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Personal Care</strong> (A)</td>
<td>1 turned on tap &amp; washed face</td>
<td>1 turned on the tap</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 combed hair</td>
<td>2 combed my hair</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 put on jewelry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 brushed teeth</td>
<td>3 prepared and brushed my teeth</td>
<td>1 prepared and brushed my teeth</td>
<td></td>
</tr>
<tr>
<td><strong>Washing</strong> (B)</td>
<td>5 got in &amp; out of tub</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 got in &amp; out of shower</td>
<td>6 got in &amp; out of the tub or shower</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 washed my whole body &amp; hair</td>
<td>7 washed my whole body</td>
<td>3 washed my whole body</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 dried my whole body &amp; hair</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Toileting</strong> (C)</td>
<td>9 got on &amp; off the toilet</td>
<td>4 got on &amp; off the toilet</td>
<td>2 used the toilet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 prepared to use the toilet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11 use the toilet</td>
<td>5 used the toilet paper &amp; flushed the toilet</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dressing</strong> (D)</td>
<td>12 got clothes ready to dress</td>
<td>8 got my clothes out</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13 put on shirt</td>
<td>9 put my shirt on</td>
<td>4 put my shirt on</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14 put on pants</td>
<td>10 put my pants on</td>
<td>5 put my pants on</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 put on socks &amp; shoes</td>
<td>12 put socks on my</td>
<td>7 put my shoes on and did them up</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 fastened buttons &amp; zippers</td>
<td>11 fastened my clothes</td>
<td>6 fastened my clothes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(buttons &amp; zippers)</td>
<td>(buttons &amp; zippers)</td>
<td></td>
</tr>
<tr>
<td><strong>Transportation</strong> (E)</td>
<td>17 got in &amp; out of a car</td>
<td>14 got in &amp; out of an automobile</td>
<td>26 got in and out of an automobile</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18 got on &amp; off a bus</td>
<td>15 got on &amp; off a bus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19 took a 1 hr. car trip</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 took a long trip</td>
<td>16 went on trips or vacations</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eating / Drinking</strong> (F)</td>
<td>21 drank from a glass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>22 ate finger foods with R hand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>23 ate finger foods with L hand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 used a knife, fork &amp; spoon</td>
<td>18 used my finger to eat small foods</td>
<td>17 ate my meals (used a knife &amp; fork)</td>
<td></td>
</tr>
<tr>
<td><strong>Food Preparation</strong> (G)</td>
<td>25 prepared a snack</td>
<td>19 made a snack</td>
<td>8 made a snack</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26 poured from a 1 L pitcher</td>
<td>20 poured myself a drink</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*continued...*
<table>
<thead>
<tr>
<th>Sub-Domain</th>
<th>ASK73</th>
<th>ASK53</th>
<th>ASK30</th>
<th>Sub-Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miscellaneous (H)</td>
<td>27 managed special medical needs</td>
<td>21 took care of my medical needs</td>
<td>10 took care of my medical needs</td>
<td>Miscellaneous</td>
</tr>
<tr>
<td>Walking / Rolling (I)</td>
<td>28 made a phone call</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>29 walked without aids</td>
<td>22 walk without any support</td>
<td>13 walk without any support</td>
<td>Locomotion</td>
</tr>
<tr>
<td></td>
<td>30 walked with aids</td>
<td>23 got around inside my home</td>
<td>14 got around inside my home</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31 walked in crowds</td>
<td>24 walked in crowds</td>
<td>15 walked in crowds</td>
<td></td>
</tr>
<tr>
<td></td>
<td>32 walked &amp; kept up with friends inside</td>
<td>25 walked indoors &amp; kept up with friends</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>33 walked/rolled about the same distance as</td>
<td>30 walked/rolled about the same distance as</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>34 walked up a hill</td>
<td>27 walked up a hill</td>
<td>17 walked up a hill</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35 walked on slippery or wet surfaces</td>
<td>28 walked on rough or slippery surfaces</td>
<td>18 walked on rough or slippery surfaces</td>
<td></td>
</tr>
<tr>
<td></td>
<td>36 walked on rough or uneven surfaces</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>37 stepped over obstacles on ground</td>
<td>26 got around outside</td>
<td>16 got around outside</td>
<td></td>
</tr>
<tr>
<td></td>
<td>38 ran/rolled &amp; kept up with friends outdoors</td>
<td>29 ran/rolled outdoors &amp; kept up with friends</td>
<td>19 ran/rolled outdoors &amp; kept up with friends</td>
<td></td>
</tr>
<tr>
<td>Carrying (J)</td>
<td>39 carried small objects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40 carried fragile or spillable objects</td>
<td>31 carried a drink or food to the table</td>
<td>20 carried a drink or food to the table</td>
<td>Standing Skills</td>
</tr>
<tr>
<td></td>
<td>41 carried objects in 2 hands</td>
<td>32 carried things in 2 hands</td>
<td>21 carried things in 2 hands</td>
<td></td>
</tr>
<tr>
<td></td>
<td>42 carried a backpack full of books</td>
<td>33 carried my backpack</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stairs (K)</td>
<td>43 got up a curb or single step</td>
<td>34 climbed up &amp; down one step or curb</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>44 got up a curb or single step</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>45 got up a full flight of stairs</td>
<td>35 climbed up &amp; down a flight of stairs</td>
<td>22 climbed up &amp; down a flight of stairs</td>
<td>Stairs</td>
</tr>
<tr>
<td></td>
<td>46 got down a full flight of stairs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47 got up a 2 flights of stairs</td>
<td>36 climbed up &amp; down 2 flights of stairs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48 climbed stairs with an alternating pattern</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49 crawled up stairs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50 crawled down stairs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>51 did quiet activities alone</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

continued ...
Table A6 Item History continued

<table>
<thead>
<tr>
<th>Sub-Domain</th>
<th>ASK73</th>
<th>ASK53</th>
<th>ASK30</th>
<th>Sub-Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spare Time</td>
<td></td>
<td></td>
<td></td>
<td>Play</td>
</tr>
<tr>
<td>Activities</td>
<td>52 did quiet activities with friends</td>
<td>37 played quiet activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(L)</td>
<td></td>
<td>38 ran and played outside with friends</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>53 did sports activities alone</td>
<td>39 played sports I enjoy by myself</td>
<td>12 played sports I enjoy by myself</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>40 participated in gym class</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>54 did sports activities with friends</td>
<td>41 played sports and active games</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sitting</td>
<td></td>
<td></td>
<td></td>
<td>Transfers</td>
</tr>
<tr>
<td>(M)</td>
<td>56 got in &amp; out of a chair</td>
<td>43 got in &amp; out of a chair</td>
<td>27 got in &amp; out of a chair</td>
<td></td>
</tr>
<tr>
<td></td>
<td>57 got in &amp; out of a wheelchair</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>58 sat on the floor without support</td>
<td>44 sat on the floor</td>
<td>28 sat on the floor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>59 sat on a chair for 1 hr</td>
<td>45 sat for at least 1 hour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N)</td>
<td>60 got in &amp; out of bed</td>
<td>46 got in &amp; out of bed</td>
<td>29 got in &amp; out of bed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>61 moved around in bed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>62 slept comfortably</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>63 sat up in bed upon waking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>64 moved around on the floor</td>
<td>48 had fun on the floor or ground</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65 got down onto the floor &amp; up again</td>
<td>47 got down onto the floor &amp; up again</td>
<td>30 got down onto the floor &amp; up again</td>
<td></td>
</tr>
<tr>
<td>Standing Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(O)</td>
<td>66 stood still for 10 minutes</td>
<td>49 stood still for 10 minutes</td>
<td>22 stood still for 10 minutes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>67 stood for 10 minutes without support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>68 stood on my tippy toes</td>
<td>50 stretched to reach a high shelf</td>
<td>23 stretched to reach a high shelf</td>
<td></td>
</tr>
<tr>
<td></td>
<td>69 picked up things off the floor</td>
<td>51 picked up things from the floor</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>70 managed door knobs &amp; locks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>71 got through light doors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entering</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exiting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(P)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>72 got through heavy doors</td>
<td>53 got through heavy doors</td>
<td>24 got through heavy doors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>73 exited the house unassisted</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX IV ASK DEVELOPMENT METHODS

Figure A3 ASK73 Development Flow Chart

- Stage I: Item Generation (chapter 2)
  - Part A
  - Part B
  - Part C: Item Reduction (chapter 3)
  - Part D

- Stage II: Item Reduction (chapter 3)
  - Part B

- Stage III: Questionnaire Formatting & Pilot Testing (chapter 4)
  - Part D

- Stage IV: Field Testing (chapter 5)
  - Part A

- Identification of items from literature
- Patient & Parent Interviews: n=20 patients, n=20 parents
- Expert Panel Meeting: n=6 experts
- Pilot Testing: n=10 patients
- Validity Testing: n=26 children
- Reliability Testing: n=18 patients
- Comparison of Capability & Performance
- Discussion of Special Problems in Paediatric Function Measurement

Subject Pool:
- 5 to 15 years old
- Musculoskeletal limitations
- Neurologically normal
- Rheumatology, rehabilitation or orthopaedic clinic
APPENDIX V  ETHICS APPROVALS & CONSENT FORMS
THE HOSPITAL FOR SICK CHILDREN
RESEARCH ETHICS BOARD

Approval & Terms of Agreement

APPLICANT/S: Ms. N. Young, Drs. J. Wright and J.I. Williams

PROJECT TITLE: “Evaluation of Paediatric Physical Disability and Exploration of Contributing Factors”

DATE OF MEETING: November 11, 1994

EXPIRY DATE: November, 1995

MEMBERS OF THE BOARD*: Dr. S. Zlotkin, Chair Dr. M. Lovett
Dr. R. Abramovitch Ms. M. McAllister
Mrs. M. P. Armstrong Ms. L. Mcleod
Mrs. B. Benoliel Dr. A. Moore
Rev. N. Headley Dr. K. Perlman
Dr. P. Joshi Ms. M. Rowell

I agree to carry out the proposed research involving human subjects in accordance with the protocol passed by the scientific peer review committee and the Research Ethics Board and using the approved consent forms (if applicable).
I shall notify the department chief and the Research Ethics Board prior to implementing any significant changes in the protocol.

SIGNATURE (INVESTIGATOR) A. Young DATE Nov 25, 1994
SIGNATURE (DEPARTMENT HEAD) DATE Nov 25, 1994

The Research Ethics Board of the Hospital for Sick Children has reviewed and approved the above-named project.

Chair, Research Ethics Board DATE Dec 1, 1994
January 3rd, 1995

ETHICAL CLEARANCE

This is to certify that a Review Committee consisting of:

Dr. D. Biggar - Chairman
Mrs. Laura Atkinson - Parent
Ms. Karen Drybrough - Clinical Nurse Specialist
Dr. D. Fehlings - Physician
Dr. Jeff Jutai - Psychology
Dr. Gilbert Miller - Physician
Dr. Cathy Steele - Statistician

has examined the proposal entitled "Evaluation of Paediatric Disability and Determination of Contributing Factors." P.I.'s: N. Young, J. L. Williams, J. G. Wright (Hospital for Sick Children), and K. K. Yoshida (U. of Toronto) including materials relating to information and consent forms, and finds it to be ethically acceptable.

Dr. D. Biggar
Chairman
CERTIFICATION OF CONSENT TO PARTICIPATE IN A RESEARCH STUDY

<table>
<thead>
<tr>
<th>1. Senior Researcher</th>
<th>Dr. D. Stanitski &amp; N.L. Young</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Project Title</td>
<td>Evaluation of Paediatric Physical Disability &amp; Exploration of Contributing Factors</td>
</tr>
<tr>
<td>3. Authorized Study Sites</td>
<td>Children's Hospital of Michigan</td>
</tr>
</tbody>
</table>

This signed and stamped document certifies that this research project and the consent form conform to the ethical standards of the Wayne State University Human Investigation Committee. To be valid this document must be firmly attached to the front of the research participant's consent form; and should be read by, or to, the research participant as part of the informed consent process.

<table>
<thead>
<tr>
<th>HIC Protocol #</th>
<th>H04-70-96(P04)-ER</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRB</td>
<td>P04</td>
</tr>
<tr>
<td>Type of Approval</td>
<td>Expedited Review</td>
</tr>
<tr>
<td>Date of Approval</td>
<td>April 19, 1996</td>
</tr>
<tr>
<td>Approval effective until</td>
<td>04-19-96 – 04-19-97</td>
</tr>
<tr>
<td>Chairperson</td>
<td>T. Constance, Pharm.D.</td>
</tr>
</tbody>
</table>

Approved APR 19 1996
WAYNE STATE UNIVERSITY PEDIATRIC INVESTIGATION COMMITTEE

Signature: [Signature]

Revised 04/21/95
C. R. C. Gibbs - Executive Director
Programme Director
J. G. Williace

Sincerely,

JWM:VK

To support this research, I agree, on behalf of Variety Village, to support this research on behalf of Variety Children. I have reviewed and understood the proposal and request for approval from the hospital for sick children and have received both scientifically and approved. I have reviewed your proposal and understood the importance of how pediatric disability affects the development of a child-report measure of disability. The development of a child-report measure of disability is an important step in improving the understanding of how pediatric disability affects the development of a child-report measure of disability. The development of a child-report measure of disability is an important step in improving the understanding of how pediatric disability affects the development of a child-report measure of disability. The development of a child-report measure of disability is an important step in improving the understanding of how pediatric disability affects the development of a child-report measure of disability.

Dear Nancy:

MSG 1x8
Toronto, Ontario
555 University Avenue
The Hospital for Sick Children
Division of Orthopaedic Surgery
Research Associate
Ms. Nancy J. Young

December 12, 1994
17th October, 1994

Ms. Nancy Young
Research Physiotherapist
The Hospital for Sick Children
555 University Avenue
Toronto, Ontario
MSG 1X8

Dear Ms. Young,

On behalf of Erinoak, I would like to express our support for your research proposal "An Evaluation of Pediatric Physical Disability and Determination of Contributing Factors." We would be pleased to participate in the study providing that we can ensure the confidentiality of clients' records. I am confident that we can develop a viable mechanism of contacting clients that will work for your research team and for Erinoak.

I am pleased to declare our support dependent upon final approval by our Board Continuous Quality Improvement Committee which meets in December 1994 and look forward to the commencement of the project.

Sincerely,

Cathy Evans
Assistant Executive Director

/wlc
RESEARCH CONSENT FORM

Title: Evaluation of Paediatric Physical Disability & Exploration of Contributing Factors

Investigators:
Nancy L. Young BScPT MSc  J. Ivan Williams PhD
Research Associate  Deputy Director of Research
Division of Orthopaedic Surgery  Institute for Clinical Evaluative Science
The Hospital for Sick Children  and Professor, University of Toronto
(416) 813-5289  (416) 480-4780

James G. Wright MD MPH FRCSC  Karen K. Yoshida PhD
Orthopaedic Surgeon  Assistant Professor
The Hospital for Sick Children  Department of Physical Therapy
and Professor, University of Toronto  University of Toronto
(416) 813-6433  (416) 978-6589

Purpose of the Research
Many children experience physical disabilities that can be improved with rehabilitation, medical or surgical treatment. At The Hospital for Sick Children we are interested in studying the effect of these treatments. Although we understand in a general sense how physical function changes following surgery or therapy, we cannot accurately measure these changes without a high quality questionnaire. We are now developing this questionnaire and would like your family’s assistance.

Description of the Research
Children between the ages of 5 and 15 who have physical disabilities and are associated with The Hospital for Sick Children, The Hugh MacMillan Rehabilitation Centre, Erin Oak or Variety Village are invited to participate in this 2-part study. In the first part of the study all participating children will be mailed 2 multiple-choice questionnaire booklets, and will be asked to complete and return the booklets by mail (postage paid by HSC). These questionnaires ask how easily and how often they have been doing common physical activities. Your child will require supervision to complete the booklets, and may also require assistance reading the questions. This process should not take more than 1.5 hours to complete. Parents will also be asked to do 2 very short questionnaires that take about 5 minutes to complete.

After completing the first part of the study, about half of the children will be asked to participate in a second part of the study. If your family is asked to join in the second part of the study, you will only be asked to do 1 of the following 3 activities:

1) come to The Hospital for Sick Children to see a physiotherapist, who will observe the child’s abilities on several activities. The assessment will not take more than 30 minutes, and will help us compare how children rate their abilities to how a therapist rates their abilities. Your family will only be asked to come in for this assessment if you live within Metro Toronto, or will be coming to The Hospital for Sick Children for another clinical appointment. You will be reimbursed for parking or TTC costs for this visit;

or

2) complete a second set of questionnaires approximately 2 to 4 months after the first set of questionnaires. This will take approximately 1 hour. This will help us see how well the questionnaires work over time. These questionnaires will also be sent to you by mail and returned by postage-paid mail. Your family will only be asked to be part of this group if your child’s physical function is expected to get noticeably better or worse over the next 6 months; or

3) participate in an interview session in your home (or at The Hospital for Sick Children if you prefer). During the interview session the parent and child are interviewed separately, and in total the session will take approximately 1 hour. We recognize that there are many reasons for
children to have restrictions in activity. Some of these are related to their diagnosis (such as weakness) and some are not. These interviews are designed to help us to learn some of the factors that affect your child’s activities. We prefer to do interviews in the home for your convenience, and to make it easier for children to identify factors that contribute to limitations in activities in the community. Your family will only be asked to be interviewed if you live within Metro Toronto, or will be here for another clinical appointment.

**Potential Harms**
There are no additional risks associated with participation in this study. For those having clinical testing: mild discomfort may be experienced during your child’s testing in the physiotherapy department. Your child will not be asked to continue if he/she asks to stop an activity due to pain.

**Potential Benefits**
The primary reason for this research is to improve our ability to measure the effect of current rehabilitation, medical and surgical treatments on disability. There is no proven benefit to your child, however, participation will permit you to assist in the development of a new questionnaire that will be used in future.

**Confidentiality**
Confidentiality will be respected and no information that discloses the identity of the subject will be released or published without consent, unless required by law. For your information, the research consent form will be inserted in the patient health record.

**Participation**
Participation in this study is voluntary. If you choose not to participate, your care and your family’s care will not be affected. If you choose on behalf of your child to participate in this study you can withdraw your child from the study at any time. Again, you and your family will continue to have access to quality care at HSC.

I acknowledge that the research procedures described above have been explained to me and that any questions that I have asked have been answered to my satisfaction. I have been informed of the alternatives to participation in this study, including the right not to participate and the right to withdraw without compromising the quality of medical care at The Hospital for Sick Children for my child and for other members of my family. As well, the potential harms and discomforts have been explained to me and I also understand the benefits (if any) of participating in the research study. I know that I may ask now, or in the future, any questions I have about the study or the research procedures. I have been assured that records relating to my child and my child’s care will be kept confidential and that no information will be released or printed that would disclose personal identity without my permission.

I hereby consent for my child ___________________________ to participate.

Our family member who may be contacted about the research is: ___________________________

Who may be reached at telephone #: ___________________________

___________________________________________

Parent's/Guardian's Name

___________________________________________

Parent's/Guardian's Signature
INFORMATION FORM FOR CHILDREN

Title of Study: Evaluation of Paediatric Physical Disability & Exploration of Contributing Factors

Investigators:
Nancy L. Young  BScPT MSc
Research Associate
Division of Orthopaedic Surgery
The Hospital for Sick Children
(416) 813-5289

James G. Wright  MD MPH FRCSC
Orthopaedic Surgeon
The Hospital for Sick Children
and Professor, University of Toronto
(416) 813-6433

J. Ivan Williams  PhD
Deputy Director of Research
Institute for Clinical Evaluative Science
and Professor, University of Toronto
(416) 480-4780

Karen K. Yoshida  PhD
Assistant Professor
Department of Physical Therapy
University of Toronto
(416) 978-6589

Why are we doing this study?
Many children have trouble doing physical activities. There are sometimes ways to help these children. We are interested in learning which treatments are best. We understand in a general way how much change in activity children can expect following surgery or therapy, but we cannot accurately measure these changes without a good questionnaire. We are now developing this questionnaire and would like your help.

What will happen during the study?
We are asking HSC patients between the ages of 5 and 15 to join our study. The study has 2 parts. In the first part we will mail two questionnaires about activities to you. You may ask someone in your family to help you read the questions. You will answer the questions and mail the package back to The Hospital for Sick Children in a special envelope.

You may later be asked to help in the second part of the study. If you are, you will only be asked to do 1 of the following 3 activities:

1) come to see a physiotherapist at The Hospital for Sick Children. He will watch you do several activities. This will take about 30 minutes and you may feel a little soreness when you are doing activities with the physiotherapist, but you may stop if you have pain;

or

2) complete a second set of questionnaires about 2 months after the first set. These will come to you by mail. After you answer the question you will send them back to us by mail;

or

3) we would like to visit with you and your family at your home (or at The Hospital for Sick Children if you like) to talk to you about things that help you with your activities, and things that make it hard for you to be active. There are many reasons why you may have problems with some
activities. Some problems may be because of your health and some may be for other reasons. We would like to learn more about this.

**Are there good things and bad things about the study?**
There are no bad things about this study, but it will take you up to 30 minutes to answer each questionnaire. The good thing about the study is that you will help make a new questionnaire that will help us learn about better treatments for kids.

**Who will know about what I did in the study?**
Nobody but your family and the staff at HSC will know you were in the study. You will be given a study number so that your name will not be written on your questionnaires. This helps us keep your answers a secret.

**Can I decide if I want to be in the study?**
You can decide if you want to be in the study or not. Your care and your family’s care will not change if you decide not to join our study. If you do join the study, it is OK to change your mind later and quit the study.

Parent to complete:

I was present when ________________ read this form and gave his/her verbal assent.

**Child’s Name**

__________________________

Name of person who obtained assent

__________________________

Signature

__________________________

Date
APPENDIX VI  GLOBAL RATING ALGORITHM

Questions Answered by Referring Clinicians

1. Is this child independent with age-appropriate self-care activities?
   - all (0)  some (1)  few (2)

2. Is this child able to get around independently?
   - unlimited (0)  limited distance (1)  only inside (2)

3. How would you rate this child's range of motion?
   - all joints normal (0)  minor limitations (1)  major limitations (2)

4. How would you rate this child's ability to participate in gym class?
   - normal (0)  limited (1)  unable or not permitted (2)

5. Does this child use a wheelchair?
   - never (0)  sometimes (1)  always (2)

Global Rating Given

Normal: Sum of all questions was zero.

Mild: Sum of questions one to four was greater than zero but less than six
   (BUT: If this was true and question five was equal to two then rated as moderate).

Moderate: Sum of questions one to four was less than six and question five equaled two.
   OR: Sum of questions one to four was greater than five but less than eight.

Severe: Sum of questions one to four was greater than seven.
**Activities Scale for Kids**

**Performance Version**

© N.L. Young

If you need help with this booklet
Call: Nancy Young
(416) 813-5289

This is a new booklet that asks you how often you did activities on your own last week.

For each question think about what you really did last week. Try to remember the things you did at home, at school, and when you were with your friends.

Record your answer by putting an X in the most correct box. If there isn't a good answer for you, write your own answer beside the question. It is important that you answer all questions.

<table>
<thead>
<tr>
<th><strong>Answers</strong></th>
<th><strong>Explanation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>all of the time</td>
<td>= I did it every time I needed to</td>
</tr>
<tr>
<td>most of the time</td>
<td>= I did it almost all of the time I needed to.</td>
</tr>
<tr>
<td></td>
<td>= but once in a while I didn't do it</td>
</tr>
<tr>
<td>sometimes</td>
<td>= I did it about half of the time I needed to.</td>
</tr>
<tr>
<td></td>
<td>= but half of the time I didn't do it</td>
</tr>
<tr>
<td>once in a while</td>
<td>= I did it at least once last week when I needed to.</td>
</tr>
<tr>
<td></td>
<td>= but most of the time I didn't do it</td>
</tr>
<tr>
<td>none of the time</td>
<td>= I didn't do it at all when I needed to</td>
</tr>
</tbody>
</table>
EXAMPLES

Terry's hands were very sore 3 days last week, so his mom helped him get dressed. He dressed himself without help the other 4 days.

One of Terry's answers looked like this . . .
I fastened my clothes by myself . . . (fastened means doing up buttons & zippers)
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
or ☐ I did not need to, because none of my clothes have fasteners

Sally's health was very good last week. She wanted a snack on 5 nights. She got a snack by herself on 4 nights.

One of Sally's answers looked like this . . .
I made a snack by myself . . . (or prepared breakfast or lunch)
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
or ☐ I am not allowed to have snacks or make my own meals

Note: If you never had a chance to do an activity last week, do not pick a box. Write your reason beside the question. For example: maybe you never came across a wet sidewalk, or maybe someone drove you to school, just to be safe.

Last week . . .

I turned on the tap by myself . . . (Examples: to wash or get a drink)
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

I combed (or brushed) my hair by myself . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
or ☐ I do not have enough hair to comb or brush

I put toothpaste on my toothbrush then brushed my teeth by myself . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

I got on and off the toilet at home by myself . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

I used the toilet paper and flushed the toilet by myself . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
Last week . . .

I got in and out of the tub or shower by myself . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

I washed my whole body by myself . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

I got my clothes out and got ready to dress by myself . . .
(Example: got clothes from my drawers or closet)
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

I put my shirt on by myself . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

I put my pants on by myself . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

I fastened my clothes by myself . . . (fastened means doing up buttons and zippers)
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

or [ ] I did not need to, because none of my clothes have fasteners

I put my socks on by myself . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

or [ ] I did not need to wear socks last week

I put my shoes on and did them up by myself . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

or [ ] I did not need to wear shoes last week

Explain:
Last week . . .

I got in and out of an automobile by myself . . .
(Opened the door, got in, closed the door, and got out again)

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

or ☐ I didn't need to go anywhere by automobile last week

I got on and off a bus by myself . . .

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

or ☐ I did not need to ride a bus last week

When there were trips or vacations planned this year I went . . .
(Examples: school bus trips or traveling to visit family in another city)

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

or ☐ There were no school trips or family trips this year

I ate my meals without help . . . (Examples: used a spoon for soup and a knife to cut food)

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

or ☐ I did not need to, because I was on a special diet

Explain:

Last week . . .

I used my fingers to eat small foods . . . (Examples: cookies, chips, smurries or raisins)

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

I made a snack by myself (or prepared breakfast or lunch) . . .

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

or ☐ I was not allowed to have snacks or make my own meals

I poured myself a drink without help . . .

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

I took care of my medical needs . . . (Examples: put on splints or took medication)

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

or ☐ I did not have special medical needs last week

I walked without any support . . . (No crutches or canes)

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
Last week...

To get around INSIDE I usually used...

☐ no special support (just my 2 feet) ☐ an artificial leg (or leg)
☐ one cane ☐ a wheelchair
☐ two canes ☐ a scooter
☐ crutches ☐ brace, splint or orthosis
☐ a walker ☐ other, please describe

Answer the questions on this page by telling how you did with the support you marked on this list.

Example 1: If you said "I usually used one cane", then answer the next questions by telling how you did with one cane last week.

Example 2: If you said "I usually used a wheelchair", then answer the next questions by telling how you did with a wheelchair last week.

I got around inside my home without anyone to help or watch me...

(Example: walked or rolled to the bathroom, to the dinner table, or to the kitchen for a drink)

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

I walked (or rode) in crowded areas...

(Example: school hallways between classes or at a busy shopping mall)

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time, because I always tried to stay away from crowds

or ☐ I did not need to, because I never came across a crowd last week

When I walked (or rode) with my friends INDOORS, I kept up with them...

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

---

Last week...

To get around OUTSIDE I usually used...

☐ no special support (just my 2 feet) ☐ an artificial leg (or leg)
☐ one cane ☐ a wheelchair
☐ two canes ☐ a scooter
☐ crutches ☐ brace, splint or orthosis
☐ a walker ☐ other, please describe

Answer the questions on this page and the next page by telling how you did with the support you marked on this list.

Example: If you said "I usually used crutches", then answer the questions that follow by telling how you did with crutches last week.

I got around outside without anyone to help or watch me...

(Example: walked or rolled to a friend's house, to school, or to the park)

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

or ☐ I never had the chance to go outside last week

Explain:

I walked (or rode) up a gentle hill or slope by myself...

☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time, because I always tried to stay away from hills

or ☐ I did not need to, because I did not come across a hill last week

---

256
Last week . . .
I walked (or ran) on rough or slippery surfaces . . .
(Examples: gravel driveways, or wet sidewalks)
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time, because I stayed away from rough and slippery surfaces
or □ I did not need to, because I did not come across a rough or slippery surface last week

When I ran (or raced) around outside, I kept up with my friends . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time
or □ I did not have a chance to be outside with my friends last week

When I walked (or raced), I usually went about the same distance as . . .
- more than twice around the outside of my school
- around the outside of my school
- the length of 2 school hallways
- the length of 1 school hallway
- the length of my classroom

The farthest distance I went (walked with support or raced) last week was . . .

The farthest distance I walked without support last week was . . .

WOW!
You are doing very well.
Take a short rest if you like.
But please try to finish this today.

How much help have you had so far?
- I have done the questions all by myself.
- Someone has read the questions to me.
- Someone has helped me with some of the answers.
- Someone has helped me with most of the answers.
- other Please describe:

Reminder:
If you never had a chance to do an activity last week, do not pick a box. Write your reason beside the question. For example: maybe you never came across a wet sidewalk, or maybe someone drove you to school, just to be safe.

I did my printing (or script writing) by myself . . .
(Examples: to do my school work)
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time
Last week . . .

I carried a drink or food to the table by myself without spilling . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
or ☐ I am not allowed to do this

I carried things in 2 hands by myself . . .
(Example: big or heavy things like stuffed animals, family pet, laundry basket)
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
or ☐ I had no reason to carry anything in 2 hands last week

I carried my backpack (or schoolbag) by myself . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
or ☐ I had no reason to carry a backpack or school bag last week

I climbed up and down one step or a curb by myself . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

Last week . . .

I walked up and down a flight of stairs (even when other people were using them) . . .
(Notes: one flight is about 14 stairs)
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
or ☐ I did not need to, because I did not come across a flight of stairs last week

I went up and down 2 flights of stairs (even when other people were using them) . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
or ☐ I did not need to, because I did not come across 2 flights of stairs last week

I did the same quiet activities that I usually do with my friends . . .
(Examples: doing puzzles with friends, playing house, board games, watching TV, playing Nintendo)
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
or ☐ I did not see my friends last week because ______________________

I ran around and played outside with my friends . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
or ☐ I was not allowed to go out last week because ______________________
Last week . . .

I did the same sports that I usually enjoy by myself . . .
(Examples: gymnastics, jungle gym, aerobics, dance, jogging, bicycling, swimming)
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

I participated in gym class (or summer camp games). . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
☐ I was not allowed to participate in gym class or summer camp games last week because of my health
☐ We didn't have gym class last week because of the holidays

I played sports and active games with friends . . .
(Examples: playing catch, basketball, T-ball, hacky-sack, tag, soccer, rough & tumble play, skimming)
Mark as many boxes as apply (for this question only)
☐ on a competitive or select team
☐ in an organized group with a set of rules (like house league)
☐ in gym class
☐ in pick-up games in park or school yard (with few rules)
☐ I was unable to play

I played these sports and active games with friends . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

Last week . . .

I did my usual job or chores . . . (Examples: paper route, baby-sitting, doing the dishes)
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
☐ I don't have a job or chores

I got in and out of a chair (or wheelchair) by myself . . .
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time

I sat on the floor . . . (Examples: at a school assembly or watching TV)
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
☐ I did not have any reason to sit on the floor

I sat for at least one hour in my usual chair at school or at the dinner table at home . . . (Example: long enough to sit through one class at school, or through dinner)
☐ all of the time
☐ most of the time
☐ sometimes
☐ once in a while
☐ none of the time
☐ I was not allowed to sit in a chair
☐ I didn't need to sit in a chair for an hour
Last week . . .

I got in and out of my bed by myself . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

I got down onto the floor or ground from standing, and got back up again by myself . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

or  I did not need to get onto the floor or ground
Explain:

I had fun on the floor or ground . . .
(Examples: played with the family pet or a wrestled with a friend, brother, or sister)
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

I stood still for 10 minutes without resting . . .
(Examples: waiting in line at the store, or talking with friends in the school hallway)
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

Last week . . .

I stretched to reach a high shelf (or to see over the person in front of me) . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

or  I did not need to stretch for anything

I picked up things from the floor by myself . . .
(Examples: bent down, squatted, or reached to pick up my knapsack or school bag from the floor)
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

I turned door knobs by myself . . .
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time

I got through heavy doors by myself . . . (Example: the front door at home or at school)
- all of the time
- most of the time
- sometimes
- once in a while
- none of the time
These questions ask how you have been feeling.

Last week my activities were:

<table>
<thead>
<tr>
<th></th>
<th>Not limited</th>
<th>A little limited</th>
<th>Somewhat limited</th>
<th>Extremely limited</th>
</tr>
</thead>
<tbody>
<tr>
<td>by my health</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Last week my freedom to move around when and how I wanted was:

<table>
<thead>
<tr>
<th></th>
<th>Not limited</th>
<th>A little limited</th>
<th>Somewhat limited</th>
<th>Extremely limited</th>
</tr>
</thead>
<tbody>
<tr>
<td>by my health</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How much of a problem did you have last week with the following…

**Pain**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A middle size problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

**Stiffness after sleeping or resting**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A middle size problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

**Limited flexibility in joints (loss of range of motion)**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A middle size problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

**Weakness**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A middle size problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

How much of a problem did you have last week with the following…

**Lack of co-ordination or balance**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A moderate problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

**Lack of energy**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A moderate problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

**Tiredness or need to take rests**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A moderate problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

**Slowness**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A moderate problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

**Not feeling safe or being told not to do certain activities**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A moderate problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

**Needing to use extra effort to do activities**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A moderate problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

**Needing to have someone else help you**

<table>
<thead>
<tr>
<th></th>
<th>No problem</th>
<th>A little problem</th>
<th>A moderate problem</th>
<th>A big problem</th>
</tr>
</thead>
</table>

YOU ARE ALL DONE  THANK YOU!!!
Please feel free to write any comments or suggestions you or your family would like to share with us in the space below.
For each question think about what you could have done last week. If you didn't try an activity last week, think about how hard it would have been to do it if you had tried.

Record your answer by putting an X in the most correct box. If there isn't a good answer for you, write your own answer beside the question. It is important that you answer all questions.

This is a new booklet that asks how you think you could have done activities last week.

Answers | Explanation
---------|--------------------------------------------------
with no problem | I could have done it with no pain, or difficulty, or extra time
with a little problem | I could have done it with a little pain, or difficulty, or extra time
with a moderate problem | I could have done it with a medium amount of pain, or difficulty, or extra time
with a big problem | I could have done it with a lot of pain, or difficulty, or extra time
I could not | I could not have done it at all
Last week, Terry's mom dressed him on 5 days, because it would have been very painful for him to do it. He dressed himself on the other 4 days because he didn't have any pain. So, over the week he had some bad days and some good days.

One of Terry's answers looked like this...

I think I could have fastened my clothes by myself...

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

Sally's health was very good last week. Every time she wanted a snack she got it by herself with a little extra effort.

One of Sally's answers looked like this...

I think I could have made a snack by myself...

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

Note: Even if you never had a chance to do an activity last week, or had no reason to do one of the activities, you still must pick a box. Imagine how hard it would have been to do it last week, and pick the best box.

Last week...

I think I could have turned on the tap by myself...

(Examples: to wash or get a drink)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have combed (or braided) my hair by myself...

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have put toothpaste on my toothbrush then brushed my teeth by myself...

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have gotten on and off the toilet at home by myself...

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have used the toilet paper and flushed the toilet by myself...

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not
Last week . . .

I think I could have gotten in and out of the tub or shower by myself . . .
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have washed my whole body by myself . . .
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have gotten my clothes out and gotten ready to dress by myself . . . (Example: got clothes from my drawers or closet)
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have put my shirt on by myself . . .
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have put my pants on by myself . . .
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

Last week . . .

I think I could have fastened my clothes by myself . . .
(fastened means doing up buttons and zippers)
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have put my socks on by myself . . .
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have put my shoes on and done them up by myself . . .
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have gotten in and out of an automobile by myself . . .
(opened the door, got in, closed the door, and got out again)
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have gotten on and off a bus by myself . . .
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not
Last week...

When there were trips or vacations planned this year I think I could have gone...

- with no problem
- with a little problem
- with a moderate problem
- with a big problem
- I could not

I think I could have eaten my meals without help...

- with no problem
- with a little problem
- with a moderate problem
- with a big problem
- I could not

I think I could have used my fingers to eat small foods...

- with no problem
- with a little problem
- with a moderate problem
- with a big problem
- I could not

I think I could have made a snack by myself (or prepared breakfast or lunch)...

- with no problem
- with a little problem
- with a moderate problem
- with a big problem
- I could not

I think I could have poured myself a drink without help...

- with no problem
- with a little problem
- with a moderate problem
- with a big problem
- I could not

Last week...

I think I could have taken care of my medical needs...

- with no problem
- with a little problem
- with a moderate problem
- with a big problem
- I could not

I think I could have walked without any support...

- with no problem
- with a little problem
- with a moderate problem
- with a big problem
- I could not

To get around INSIDE I usually used...

- no special support (just my 2 feet)
- an artificial leg (or legs)
- one cane
- a wheelchair
- two canes
- a scooter
- crutches
- a brace, splint or orthosis
- a walker
- other, please describe

Answer the questions on this page and the next page by telling how you could have done with the support you marked on this list.

Example 1: If you said "I usually used one cane", then answer the next questions by telling how you could have done with one cane last week.

Example 2: If you said "I usually used a wheelchair", then answer the next questions by telling how you could have done with a wheelchair last week.

I think I could have gotten around inside my home without anyone to help or watch me...

- with no problem
- with a little problem
- with a moderate problem
- with a big problem
- I could not
Last week . . .

I think I could have walked (or rode) in crowded areas . . .
(Example: school hallways between classes or at a busy shopping mall)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not, because I always tried to stay away from crowds

When I walked (or rode) with my friends INDOORS, I think I could have kept up with them . . .

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

To get around OUTSIDE I usually used . . .

☐ no special support (just my 2 feet) ☐ an artificial leg (or leg)
☐ one cane ☐ a wheelchair
☐ two canes ☐ a scooter
☐ crutches ☐ brace, splint or orthosis
☐ a walker ☐ other, please describe

Answer the questions on this page and the next page by telling how you could have done with the support you marked on this list.

Example: If you said "I usually used crutches", then answer the questions that follow by telling how you could have done with crutches last week.

I think I could have gotten around outside without anyone to help or watch me . . .
(Example: walked or rode to a friend's house, to school, or to the park)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

Last week . . .

I think I could have walked (or rode) up a gentle hill or slope by myself . . .

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not, because I always tried to stay away from hills

I think I could have walked (or rode) on rough or slippery surfaces . . .
(Examples: gravel driveways, or wet sidewalks)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not, because I stayed away from rough and slippery surfaces

When I ran (or rode) around outside, I think I could have kept up with my friends . . .

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

When I walked (or rolled), I think I could have usually gone about the same distance as . . .

☐ more than twice around the outside of my school
☐ around the outside of my school
☐ the length of 2 school hallways
☐ the length of 1 school hallway
☐ the length of my classroom

The farthest distance I think I could have gone (walked with support or rolled) last week was . . .

☐ in the blank

The farthest distance I think I could have walked without support last week was . . .

☐ in the blank

267
WOW!
YOU ARE DOING VERY WELL.
TAKE A SHORT REST IF YOU LIKE.
BUT PLEASE TRY TO FINISH THIS TODAY.

HOW MUCH HELP HAVE YOU HAD SO FAR?
☐ I have done the questions all by myself.
☐ Someone has read the questions to me.
☐ Someone has helped me with some of the answers.
☐ Someone has helped me with most of the answers.
☐ other  Please describe:

Reminder:
Even if you never had a chance to do an activity last week, or had no reason to do one of the activities, you still must pick a box. Imagine how hard it would have been to do it last week, and pick the best box...

I think I could have carried a drink or food to the table by myself without spilling...
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have carried things in 2 hands by myself...
(Example: big or heavy things like stuffed animals, family pet, laundry basket)
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have carried my backpack (or school bag) by myself...
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have climbed up and down one step or a curb by myself...
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have walked up and down a flight of stairs (even when other people were using them)...
(Note: one flight is about 14 stairs)
☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not
I think I could have gone up and down 2 flights of stairs (even when other people were using them) . . .

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have done the same quiet activities that I usually do with my friends . . . (Doing puzzles with friends, playing house, board games, watching TV, playing Nintendo)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have run around and played outside with my friends . . .

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have done the same sports that I usually enjoy by myself . . .

(Examples: gymnastics, jungle gym, aerobics, dance, jogging, bicycling, swimming)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have participated in gym class (or summer camp games) . . .

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have played sports and active games with friends . . .

(Examples: playing catch, basketball, T-ball, hocky-sac, tag, soccer, rough & tumble play, skiing)

Mark as many boxes as apply (for this question only)

☐ on a competitive or select team
☐ in an organized group with a set of rules (like house league)
☐ in gym class
☐ in pickup games in park or school yard (with few rules)
☐ I was unable to play

I think I could have done my usual job or chores . . .

(Examples: paper route, baby-sitting, doing the dishes)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have gotten in and out of a chair (or wheelchair) by myself . . .

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have sat on the floor . . . (Examples: at a school assembly or watching TV)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not
I sat for at least one hour in my usual chair at school or at the dinner table at home... (Example: long enough to sit through one class at school, or through dinner)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have gotten in and out of my bed by myself...

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have gotten down onto the floor or ground from standing, and gotten back up again by myself...

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have had fun on the floor or ground...

(Examples: played with the family pet or wrestled with a friend, brother, or sister)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have stood still for 10 minutes without resting...

(Examples: waiting in line at the store, or talking with friends in the school hallway)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have stretched to reach a high shelf (so to see over the person in front of me)...

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have picked up things from the floor by myself...

(Examples: bent down, squatted, or reached to pick up my knapsack or school bag from the floor)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have turned door knobs by myself...

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

I think I could have gotten through heavy doors by myself...

(Example: the front door at home or at school)

☐ with no problem
☐ with a little problem
☐ with a moderate problem
☐ with a big problem
☐ I could not

YOU ARE ALL DONE

THANK YOU!!!
## APPENDIX VIII  HEALTH UTILITIES MARK III QUESTIONNAIRE

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Level</th>
<th>Level description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision</td>
<td>1</td>
<td>Able to see well enough to read ordinary newspaper and recognize a friend on the other side of the street, without glasses or contact lenses.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Able to see well enough to read ordinary newspaper and recognize a friend on the other side of the street, but with glasses or contact lenses.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Able to read ordinary newspaper with or without glasses but unable to recognize a friend on the other side of the street, even with glasses or contact lenses.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Able to recognize a friend on the other side of the street with or without glasses but unable to read ordinary newspaper, even with glasses or contact lenses.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Unable to read ordinary newspaper and unable to recognize a friend on the other side of the street, even with glasses or contact lenses.</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Unable to see at all.</td>
</tr>
<tr>
<td>Hearing</td>
<td>1</td>
<td>Able to hear what is said in a group conversation with at least 3 other people, without a hearing aid.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Able to hear what is said in a conversation with 1 other person in a quiet room without a hearing aid, but requires a hearing aid to hear what is said in a group conversation with at least 3 other people.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Able to hear what is said in a conversation with 1 other person in a quiet room with a hearing aid, and able to hear what is said in a group conversation with at least 3 other people, with a hearing aid.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Able to hear what is said in a conversation with 1 other person in a quiet room without a hearing aid, but unable to hear what is said in a group conversation with at least 3 other people, even with a hearing aid.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Able to hear what is said in a conversation with 1 other person in a quiet room with a hearing aid, but unable to hear what is said in a group conversation with at least 3 other people even with a hearing aid.</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Unable to hear at all.</td>
</tr>
<tr>
<td>Speech</td>
<td>1</td>
<td>Able to be understood completely when speaking with strangers or people who know me well.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Able to be understood partially when speaking with strangers but able to be understood completely when speaking with people who know me well.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Able to be understood partially when speaking with strangers or people who know me well.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Unable to be understood when speaking with strangers but able to be understood partially by people who know me well.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Unable to be understood when speaking to other people (or unable to speak at all).</td>
</tr>
<tr>
<td>Ambulation</td>
<td>1</td>
<td>Able to walk around the neighborhood without difficulty, and without walking equipment.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Able to walk around the neighborhood with difficulty, but does not require walking equipment or the help of another person.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Able to walk around the neighborhood with walking equipment, but without the help of another person.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Able to walk only short distances with walking equipment, and requires a wheelchair to get around the neighborhood.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Unable to walk alone, even with walking equipment. Able to walk short distances with the help of another person, and requires a wheelchair to get around the neighborhood.</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Cannot walk at all.</td>
</tr>
<tr>
<td>Dexterity</td>
<td>1</td>
<td>Full use of 2 hands and 10 fingers.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Limitations in the use of hands or fingers, but does not require special tools or help of another person.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Limitations in the use of hands or fingers, is independent with use of special tools and does not require the help of another person.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Limitations in the use of hands or fingers, requires the help of another person for some tasks (not independent even with use of special tools).</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Limitations in the use of hands or fingers, requires the help of another person for most tasks (not independent even with use of special tools).</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Limitations in the use of hands or fingers, requires the help of another person for all tasks (not independent even with use of special tools).</td>
</tr>
<tr>
<td>Emotion</td>
<td>1</td>
<td>Happy and interested in life.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Somewhat happy.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Somewhat unhappy.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Very unhappy.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>So unhappy that life is not worthwhile.</td>
</tr>
<tr>
<td>Cognition</td>
<td>1</td>
<td>Able to remember most things, think clearly and solve day to day problems.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Able to remember most things, but has a little difficulty when trying to think and solve day to day problems.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Somewhat forgetful, but able to think clearly and solve day to day problems.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Somewhat forgetful, and has a little difficulty when trying to think or solve day to day problems.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Very forgetful, and has great difficulty when trying to think or solve day to day problems.</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Unable to remember anything at all, and unable to think or solve day to day problems.</td>
</tr>
<tr>
<td>Pain</td>
<td>1</td>
<td>Free of pain and discomfort.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Mild to moderate pain that prevents no activities.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Moderate pain that prevents a few activities.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Moderate to severe pain that prevents some activities.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Severe pain that prevents most activities.</td>
</tr>
</tbody>
</table>

Reproduced by permission from Table III of "Multi-Attribute Health Status Classification Systems. Health Utilities Index" [Feeny et al. 1995].
### Table A7 Summary Score Calculations

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item Scores</th>
<th>Aggregation</th>
<th>Penalty Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASKp</td>
<td>0 to 4</td>
<td>$100 \times \left( \frac{\sum \text{item responses}}{# \text{of items answered} \times 4} \right)$</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>0 = none of the time</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 = once in a while</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = sometimes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 = most of the time</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 = all of the time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASKc</td>
<td>0 to 4</td>
<td>$100 \times \left( \frac{\sum \text{item responses}}{# \text{of items answered} \times 4} \right)$</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>0 = I could not</td>
<td>usually 53 items</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 = with a big problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = with a moderate problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 = with a little problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 = with no problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASK observational</td>
<td>0 to 3</td>
<td>$100 \times \left( 1 - \frac{\sum \text{item responses}}{# \text{of items answered} \times 3} \right)$</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>0 = I could not</td>
<td>usually 15 of 30 items</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 = with a big problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = with a little problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 = with no problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHAQ†</td>
<td>0 to 3</td>
<td>$100 \times \left( 1 - \frac{\sum \text{max. sub-domain scores}^*}{# \text{of sub-domains} \times 3} \right)$</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>0 = without any difficulty</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 = with some difficulty</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = with much difficulty</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 = unable to do</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUI3 domains</td>
<td>0 to 8</td>
<td>number of domains with limitations</td>
<td>none</td>
</tr>
<tr>
<td>HUI3 ambulation sub-domain</td>
<td>1 to 6 where 6=unable</td>
<td>raw score based on description [Boyle et al. 1995, Feeny et al. 1995]</td>
<td>none</td>
</tr>
</tbody>
</table>

† Scoring has been reversed from usual presentation so that this scale also has a range of 0 to 100, where 100 = best function.
APPENDIX X  ITEM LOGITS IN QUESTIONNAIRE ORDER

Table A8 Repeat of Table 10 (ASKp53 Rasch Model) In Item Entry Order

<table>
<thead>
<tr>
<th>Item Description</th>
<th>Logit</th>
<th>Error</th>
<th>Infit</th>
<th>Outfit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 turned on the tap</td>
<td>0.61</td>
<td>0.13</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>2 combed my hair</td>
<td>0.00</td>
<td>0.13</td>
<td>1.06</td>
<td>1.43</td>
</tr>
<tr>
<td>3 prepared and brushed my teeth</td>
<td>0.61</td>
<td>0.13</td>
<td>0.80</td>
<td>0.82</td>
</tr>
<tr>
<td>4 got on &amp; off the toilet</td>
<td>0.44</td>
<td>0.13</td>
<td>0.94</td>
<td>0.68</td>
</tr>
<tr>
<td>5 used the toilet paper &amp; flushed the toilet</td>
<td>0.49</td>
<td>0.13</td>
<td>1.04</td>
<td>0.81</td>
</tr>
<tr>
<td>6 got in &amp; out of the tub or shower</td>
<td>-0.55</td>
<td>0.12</td>
<td>0.78</td>
<td>0.69</td>
</tr>
<tr>
<td>7 washed my whole body</td>
<td>0.17</td>
<td>0.12</td>
<td>0.93</td>
<td>1.04</td>
</tr>
<tr>
<td>8 got my clothes out</td>
<td>-0.56</td>
<td>0.12</td>
<td>0.64</td>
<td>0.61</td>
</tr>
<tr>
<td>9 put my shirt on</td>
<td>0.78</td>
<td>0.14</td>
<td>0.86</td>
<td>0.77</td>
</tr>
<tr>
<td>10 put my pants on</td>
<td>0.09</td>
<td>0.12</td>
<td>0.71</td>
<td>0.63</td>
</tr>
<tr>
<td>11 fastened my clothes (buttons &amp; zippers)</td>
<td>0.06</td>
<td>0.14</td>
<td>0.91</td>
<td>0.85</td>
</tr>
<tr>
<td>12 put socks on my</td>
<td>-0.08</td>
<td>0.13</td>
<td>0.75</td>
<td>0.67</td>
</tr>
<tr>
<td>13 put my shoes on and did them up</td>
<td>-0.47</td>
<td>0.12</td>
<td>0.98</td>
<td>0.94</td>
</tr>
<tr>
<td>14 got in &amp; out of an automobile</td>
<td>-0.55</td>
<td>0.12</td>
<td>0.75</td>
<td>0.64</td>
</tr>
<tr>
<td>15 got on &amp; off a bus</td>
<td>-0.51</td>
<td>0.18</td>
<td>1.10</td>
<td>0.96</td>
</tr>
<tr>
<td>16 went on trips or vacations</td>
<td>1.16</td>
<td>0.15</td>
<td>1.60</td>
<td>2.64</td>
</tr>
<tr>
<td>17 ate my meals (used a knife &amp; fork)</td>
<td>1.25</td>
<td>0.15</td>
<td>1.10</td>
<td>1.23</td>
</tr>
<tr>
<td>18 used my finger to eat small foods</td>
<td>2.35</td>
<td>0.21</td>
<td>1.52</td>
<td>1.89</td>
</tr>
<tr>
<td>19 made a snack</td>
<td>-0.93</td>
<td>0.12</td>
<td>0.70</td>
<td>0.66</td>
</tr>
<tr>
<td>20 poured myself a drink</td>
<td>-0.39</td>
<td>0.12</td>
<td>0.77</td>
<td>0.78</td>
</tr>
<tr>
<td>21 took care of my medical needs</td>
<td>-0.34</td>
<td>0.14</td>
<td>1.24</td>
<td>1.32</td>
</tr>
<tr>
<td>22 walk without any support</td>
<td>-0.52</td>
<td>0.12</td>
<td>1.25</td>
<td>1.32</td>
</tr>
<tr>
<td>23 got around inside my home</td>
<td>1.21</td>
<td>0.15</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>24 walked in crowds</td>
<td>0.46</td>
<td>0.14</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>25 walked indoors &amp; kept up with friends</td>
<td>-0.01</td>
<td>0.13</td>
<td>1.03</td>
<td>1.32</td>
</tr>
<tr>
<td>26 got around outside</td>
<td>-0.05</td>
<td>0.13</td>
<td>1.12</td>
<td>0.98</td>
</tr>
<tr>
<td>27 walked up a hill</td>
<td>0.17</td>
<td>0.15</td>
<td>0.95</td>
<td>0.80</td>
</tr>
<tr>
<td>28 walked on rough or slippery surfaces</td>
<td>-0.15</td>
<td>0.14</td>
<td>1.00</td>
<td>0.90</td>
</tr>
<tr>
<td>29 ran/rolled outdoors &amp; kept up with friends</td>
<td>-0.80</td>
<td>0.13</td>
<td>1.27</td>
<td>1.57</td>
</tr>
<tr>
<td>30 walked/rolled about the same distance as</td>
<td>-0.11</td>
<td>0.13</td>
<td>1.00</td>
<td>1.12</td>
</tr>
</tbody>
</table>

continued...
Table A8  ASKp3 Rasch Model — In Item Entry Order  

<table>
<thead>
<tr>
<th>Item Description</th>
<th>Logit</th>
<th>Error</th>
<th>Infit</th>
<th>Outfit</th>
</tr>
</thead>
<tbody>
<tr>
<td>31 carried a drink or food to the table</td>
<td>-0.23</td>
<td>0.12</td>
<td>0.80</td>
<td>0.81</td>
</tr>
<tr>
<td>32 carried things in 2 hands</td>
<td>-0.53</td>
<td>0.12</td>
<td>0.91</td>
<td>0.84</td>
</tr>
<tr>
<td>33 carried my backpack</td>
<td>0.27</td>
<td>0.15</td>
<td>1.14</td>
<td>1.02</td>
</tr>
<tr>
<td>34 climbed up &amp; down one step or curb</td>
<td>0.11</td>
<td>0.12</td>
<td>0.90</td>
<td>0.75</td>
</tr>
<tr>
<td>35 climbed up &amp; down a flight of stairs</td>
<td>-0.29</td>
<td>0.13</td>
<td>0.92</td>
<td>0.76</td>
</tr>
<tr>
<td>36 climbed up &amp; down 2 flights of stairs</td>
<td>-0.67</td>
<td>0.15</td>
<td>0.83</td>
<td>0.72</td>
</tr>
<tr>
<td>37 played quiet activities</td>
<td>0.79</td>
<td>0.14</td>
<td>1.08</td>
<td>1.31</td>
</tr>
<tr>
<td>38 ran and played outside with friends</td>
<td>-0.65</td>
<td>0.12</td>
<td>1.37</td>
<td>1.36</td>
</tr>
<tr>
<td>39 played sports I enjoy by myself</td>
<td>-0.93</td>
<td>0.12</td>
<td>1.38</td>
<td>1.49</td>
</tr>
<tr>
<td>40 participated in gym class</td>
<td>-0.69</td>
<td>0.15</td>
<td>2.03</td>
<td>2.57</td>
</tr>
<tr>
<td>41 played sports and active games</td>
<td>-1.26</td>
<td>0.13</td>
<td>1.51</td>
<td>1.82</td>
</tr>
<tr>
<td>42 did my usual job or chores</td>
<td>-0.74</td>
<td>0.15</td>
<td>0.74</td>
<td>0.84</td>
</tr>
<tr>
<td>43 got in &amp; out of a chair</td>
<td>0.74</td>
<td>0.14</td>
<td>0.77</td>
<td>0.53</td>
</tr>
<tr>
<td>44 sat on the floor</td>
<td>0.19</td>
<td>0.13</td>
<td>1.02</td>
<td>1.21</td>
</tr>
<tr>
<td>45 sat for at least 1 hour</td>
<td>0.89</td>
<td>0.15</td>
<td>1.41</td>
<td>1.83</td>
</tr>
<tr>
<td>46 got in &amp; out of bed</td>
<td>0.58</td>
<td>0.13</td>
<td>0.79</td>
<td>0.57</td>
</tr>
<tr>
<td>47 got down onto the floor &amp; up again</td>
<td>-0.04</td>
<td>0.12</td>
<td>0.99</td>
<td>0.82</td>
</tr>
<tr>
<td>48 had fun on the floor or ground</td>
<td>-0.06</td>
<td>0.12</td>
<td>1.56</td>
<td>1.61</td>
</tr>
<tr>
<td>49 stood still for 10 minutes</td>
<td>-1.06</td>
<td>0.12</td>
<td>1.04</td>
<td>1.03</td>
</tr>
<tr>
<td>50 stretched to reach a high shelf</td>
<td>-0.55</td>
<td>0.15</td>
<td>0.70</td>
<td>0.81</td>
</tr>
<tr>
<td>51 picked up things from the floor</td>
<td>0.28</td>
<td>0.13</td>
<td>0.70</td>
<td>0.66</td>
</tr>
<tr>
<td>52 turned door knobs</td>
<td>0.63</td>
<td>0.13</td>
<td>0.80</td>
<td>0.79</td>
</tr>
<tr>
<td>53 got through heavy doors</td>
<td>-0.60</td>
<td>0.12</td>
<td>0.67</td>
<td>0.66</td>
</tr>
</tbody>
</table>