

Occupational performance in school settings

Evaluation and intervention using
the School AMPS

Michaela Munkholm



**Department of Community Medicine and Rehabilitation,
Occupational Therapy, Umeå University, Sweden**
Umeå 2010

Responsible publisher under Swedish law: The Dean of the Medical Faculty
This work is protected by the Swedish Copyright Legislation (Act 1960:729)
ISBN: 978-91-7459-121-7
ISSN: 0346-6612
Copyright©Michaela Munkholm
Cover photo & design: Michael Munkholm & Casper Munkholm
E-version available at <http://umu.diva-portal.org/>
Printed by Arkitektkopia, Umeå, Sweden, 2010

Best practice is a way of thinking about problems in imaginative ways, applying knowledge creatively to solve performance problems, and taking responsibility for evaluating the effectiveness of the innovations to inform future practices. Remember, what is best practice today is out of date in the future.

(Dunn, 2000, p. 2)

Table of contents

Abstract	vi
Svensk sammanfattning (Summary in Swedish)	viii
Abbreviations	xi
Original papers	xii
Preface	xiii
Introduction	1
Occupation, doing, and occupational performance	1
<i>Areas of occupation</i>	2
<i>Levels of doing</i>	3
<i>Occupational participation</i>	3
<i>Occupational performance</i>	3
<i>Occupational skills</i>	4
Childhood roles and the interaction between occupation and roles expectations	5
Students with mild disabilities or who are at risk	6
<i>Reasons for referral to occupational therapy</i>	8
Occupational therapy assessment and interventions in school settings	8
<i>Occupational therapy as a related service</i>	8
<i>Ecologically relevant and occupation-based occupational therapy services</i>	9
<i>Top-down versus bottom-up approach to assessment</i>	10
<i>School Version of the Assessment of Motor and Process Skills (School AMPS)</i>	12
Development of the School AMPS – issues of reliability and validity	15
<i>Development of instruments using Rasch models</i>	15
<i>The many-facet Rasch model for the School AMPS</i>	16
<i>Aspects of reliability</i>	16
<i>Aspects of validity</i>	18
<i>Differential item functioning</i>	19
<i>Differential test functioning</i>	21
<i>Sensitivity of measures</i>	21
<i>Previous evidence for reliability and validity of the School AMPS</i>	22
Aims of this Thesis	24
Specific research aims	24
Methods	26
Participants	26
<i>Data extraction procedures</i>	28
<i>Study 1</i>	29

<i>Study II</i>	29
<i>Study III</i>	29
<i>Study IV</i>	30
Instrumentation	33
Procedures and data analyses	34
<i>Many-faceted Rasch analyses</i>	34
<i>Descriptive and inferential statistical analyses</i>	36
<i>Study I</i>	37
<i>Evidence of reliability</i>	37
<i>Study II</i>	39
<i>Evidence of validity based on internal structure</i>	39
<i>Evidence of validity related to consequences of testing (DTF)</i>	40
<i>Study III</i>	42
<i>Evidence of validity based relations to other variables (sensitivity)</i>	42
<i>Evidence of validity based on internal structure (DIF)</i>	42
<i>Evidence of validity based on consequences of testing</i>	43
<i>Study IV</i>	43
<i>Evidence of validity based on relations to other variables (sensitivity)</i>	43
<i>Sample size and effect size considerations</i>	44
Ethical considerations	44
Results	46
Evidence of reliability and validity of the School AMPS measures and scales	46
<i>Evidence of reliability</i>	46
<i>Evidence of validity based on internal structure and consequences of testing</i>	47
<i>Stability of the skill item hierarchies (DIF)</i>	47
<i>Differential test functioning (DTF)</i>	50
<i>Identification of actual differences in skill challenge</i>	51
<i>Evidence based on relations to other variables</i>	52
<i>Sensitivity when discriminating between typically-developing students and students with mild disabilities</i>	52
<i>Sensitivity when measuring change before and after intervention</i>	52
<i>Description of occupational therapy services</i>	57
Discussion	58
<i>Evidence of reliability</i>	59
<i>Evidence of validity based on internal structure and consequences of testing</i>	62
<i>Stability of the skill items (DIF)</i>	62
<i>Differential test functioning (DTF)</i>	64
<i>Identification of actual differences in skill challenge</i>	65
<i>Evidence based on relations to other variables</i>	66
Other methodological considerations	72
Summary of clinical implications	73

Conclusions	75
Acknowledgements	77
References	81

Abstract

Background: This thesis is comprised of four studies designed to evaluate aspects of reliability and validity of the School Version of the Assessment of Motor and Process Skills (School AMPS) (Fisher, Bryze, Hume, & Griswold, 2007), an observation-based evaluation of quality of occupational performance when children perform schoolwork tasks in school settings. The long term goal was to contribute to knowledge about children at risk or with mild disabilities who experience difficulties with occupational performance in school settings, and describe how the School AMPS can be used when a true top-down process of planning and implementing school-based occupational therapy services is implemented in a Swedish context.

Methods: In Study I, two different split-half methods and Pearson product moment correlation were used to estimate reliability of the School AMPS measures. These were cross-validated using Rasch equivalent of Cronbach's alpha. The standard error of measurement (*SEM*) was also calculated. In Studies II and III, many-facet Rasch analyses and/or relevant inferential statistics (e.g., ANOVA, *t* tests) were used to examine for evidence of validity based on (1) internal structure related to differential item functioning (DIF), (2) relations to other variables (sensitivity) in terms of comparing groups (typically-developing children vs. children with mild disabilities), and (3) consequences of testing (benefits of testing) in terms of test fairness. In Study IV, ANOVA and *t* tests were used to examine relations to other variables in terms of sensitivity of the School AMPS measures for detecting change based on repeated School AMPS evaluations pre- and post-interventions.

Results: The three methods for estimating reliability of the School AMPS measures yielded high reliability coefficient estimates ($r \geq 0.73$) and low *SEMs*. Minimal DIF was identified, and despite minimal DIF, the School AMPS measures were found to be free of differential test functioning. The School AMPS measures were sensitive enough to detect differences between groups as well as changes following consultative occupational therapy services provided in natural school settings.

Conclusions: The results support the reliability and validity of the School AMPS scales and measures when used to evaluate quality of occupational performance in school settings. The results from this thesis are of clinical importance as they provide evidence that occupational therapists can have confidence in the School AMPS measures when they are used in the process of making decisions about individual students, planning interventions, and later perform follow-up evaluations to measure the outcomes. We also have objective evidence that children with mild disabilities demonstrate diminished quality of “doing” when performing schoolwork tasks. The potential long term benefits of such evidence may be to support or justify the need for children with mild disabilities to receive occupational therapy services within school settings in Sweden; and through collaboration with teachers, plan and implement better targeted and more effective interventions.

Key words: Occupational therapy, Assessment, Rasch measurement, Children, Schoolwork performance, Differential item functioning, Differential test functioning, Outcomes, Efficacy of intervention, School-based practice, Instrument development

Svensk sammanfattning (Summary in Swedish)

Bakgrund: Avhandlingen består av fyra studier som avser att utvärdera reliabilitet och validitet hos bedömningsinstrumentet Skol-AMPS (*School Version of the Assessment of Motor and Process Skills* [Skol-AMPS]) (Fisher, et al., 2007). Skol-AMPS är en observationsbaserad bedömning som arbetsterapeuter använder för att bedöma kvalitén i barns utförande av skoluppgifter i klassrumsmiljö, planera mål och åtgärder samt mäta förändring. Det långsiktiga målet var att bidra med kunskap om barn med lindriga funktionshinder som ofta uppvisar svårigheter i utförandet av skoluppgifter, till exempel, på grund av svårigheter att koncentrera sig, motorisk oro, svårigheter att följa instruktioner eller organisera och planera sitt skolarbete för att få det färdigt i tid. Vi avsåg också att utvärdera om Skol-AMPS är ett användbart bedömningsinstrument för att bedöma barn när de utför skoluppgifter som de vill och behöver kunna utföra för att vara delaktiga i skolan och beskriva hur Skol-AMPS kan användas tillsammans med lärare eller annan skolpersonal för att planera mål och genomföra åtgärder som underlättar utförandet av skoluppgifter för barn med lindriga funktionshinder i svensk skolmiljö.

När Skol-AMPS används i det kliniska arbetet är det viktigt att kvalitetsmått är stabila. Detta för att det ska vara möjligt att mäta förändring efter insatta åtgärder. Eftersom Skol-AMPS är ett internationellt standardiserat bedömningsinstrument som på senare år har börjat användas i de nordiska länderna, framför allt i Sverige och Danmark, är det också viktigt att undersöka om Skol-AMPS *skill items* skiljer sig åt mellan olika världs regioner eller mellan olika grupper av barn. Normalt finns det små variationer men om variationerna blir för stora föreligger det risk för *differential item functioning* (DIF). Om det finns DIF mellan världs regioner kan det leda till att, till exempel, barn som bedöms med Skol-AMPS i en världs region inte får tillgång till åtgärder som det annars skulle ha fått om samma barn var bedömt i en annan världs region – bedömningen blir orättvis. Därför är det också viktigt att undersöka om DIF leder till *differential test functioning* (DTF). Om Skol-AMPS uppvisar kvalitetsmått som är reliabla och valida kan

arbetsterapeuter vara säkra på att Skol-AMPS kvalitetsmått är känsliga för att mäta förändringar över tid.

Metoder: I studie I användes två olika *split-half* metoder och *Pearson product moment correlation* för att undersöka om Skol-AMPS kvalitetsmått är stabila mellan två bedömningar. Resultatet från de båda *split-half* metoderna korsvaliderades med *Cronbach's alphas* motsvarighet i Rasch. Vi beräknade också *standard error of measurement (SEM)* som kan användas för att mäta förändring för ett enskilt barn. I studie II och III användes mångfacetterade Rasch analyser och/eller andra relevanta statistiska metoder (t.ex. ANOVA, *t* tests) för att undersöka evidens för validitet baserat på (1) intern struktur (stabilitet) relaterat till DIF, (2) relationer till andra variabler (sensivitet/känslighet) vid jämförelser mellan grupper (typiskt utvecklade barn jmf med barn med lindriga funktionshinder) och (3) konsekvenser av testning (fördelar) relaterat till om bedömningen är rättvis. I studie IV användes ANOVA och *t* tester för att undersöka om Skol-AMPS kvalitetsmått är en känsliga och kan användas för att upptäcka förändringar över tid, till exempel, före och efter insatta åtgärder.

Resultat: De tre metoderna för att beräkna reliabiliteten i Skol-AMPS kvalitetsmått ledde till höga reliabilitetskoefficienter ($r \geq 0.73$) och låga SEM. Minimal DIF identifierades mellan världs regioner och mellan typisk utvecklade barn och barn med milda funktionshinder. Trots minimal DIF visade det sig att Skol-AMPS bedömningarna var utan DTF. Resultatet visade också att Skol-AMPS är känslig för att upptäcka skillnader mellan grupper liksom mäta förändringar till följd av åtgärder som planerats tillsammans med läraren och som implementerats i barnets naturliga skolmiljö.

Sammanfattning: Resultaten bekräftar god reliabilitet och validitet hos Skol-AMPS kvalitetsmått och *skill items* när bedömningsinstrumentet används för att bedöma kvalitén i utförandet av uppgifter i skolan. Resultaten har också klinisk betydelse eftersom de ger evidens för att arbetsterapeuter kan använda Skol-AMPS med god tillförlitlighet när bedömningsresultat används för att fatta beslut om enskilda barn, planera mål och åtgärder tillsammans med läraren samt senare utföra uppföljande bedömningar för att mäta effekten av dessa åtgärder. Vi har också evidens att barn med lindriga funktionshinder faktiskt uppvisar

nedsatt kvalité i "görandet" när de utför sina skoluppgifter. De potentiella långsiktiga fördelarna med dessa evidens kan vara att motivera behovet av stöd till barn med lindriga funktionshinder som ofta faller mellan stolarna när resurser i skolan prioriteras, att de får tillgång till arbetsterapeutisk bedömning och insatser i skolan samt genom att samarbeta med lärare och föräldrar, planera och utföra bättre målinriktade och mer effektiva åtgärder implementerade direkt i barnets skolmiljö. Det återstår att genomföra fördjupade forskningsstudier, dvs. fortsätta processen att inhämta evidens om effekten av arbetsterapi i skolan och att vidareutveckla samt utvärdera Skol-AMPS användbarhet för att bedöma barn med aktivitetsbegränsningar i sin skolmiljö.

Abbreviations

ADHD	Attention deficit hyperactivity disorder
ADL	Activities of daily living
AMPS	Assessment of Motor and Process Skills
ANZ	Australia and New Zealand
CSA	Central and South America and Caribbean
D_i	Item difficulty calibration
DCD	Developmental coordination disorder
DD	Developmental disabilities
DIF	Differential item functioning
DTF	Differential test functioning
ETS	Educational Testing Services
IADL	Instrumental activities of daily living
IDEA	Individuals with Disabilities Education Act
LD	Learning disabilities
M	Mean
MFR	Many-faceted Rasch
NA	North America (United States and Canada)
OT	Occupational therapy
OTIPM	Occupational Therapy Intervention Process Model
SD	Standard deviation
SEm	Standard error of measurement
School AMPS	School Version of the Assessment of Motor and Process Skills
SFA	School Function Assessment
SI	Sensory integrative dysfunction
SqRt	Square root
UK	United Kingdom and the Republic of Ireland
WHO	World Health Organization

Original papers

This thesis is based on the following papers:

- I Munkholm, M., Löfgren, B., & Fisher, A. G. (2010). Reliability of the School AMPS measures. *Scandinavian Journal of Occupational Therapy, Early online*, 1-7.
doi:10.3109/11038128.2010.525721
- II Munkholm, M., Berg, B., Löfgren, B., & Fisher, A. G. (2010). Cross-regional validation of the School Version of the Assessment of Motor and Process Skills. *American Journal of Occupational Therapy*, 64, 768 –775.
- III Munkholm, M., & Fisher, A. G. (2008). Differences in schoolwork performance between typically-developing students and students with mild disabilities. *OTJR: Occupation, Participation, and Health*, 3, 121–132.
- IV Munkholm, M., Löfgren, B., & Fisher, A. G. (manuscript). Improving quality of schoolwork task performance through school-based consultative occupational therapy services.

Original papers have been reproduced with kind permission from the publishers.

Preface

I was still an occupational therapy student when my son entered the first grade, and, as a parent, I enjoyed spending time in his classroom. I also saw some children struggle with performing routine schoolwork tasks. They had difficulties following instructions, organizing their workspaces, concentrating on their schoolwork and getting it done on time, staying in their seats, coping with classroom rules, interacting with peers, etc. This raised many questions about the lack of support for children without diagnosed disabilities, but who clearly have occupational performance problems in their school environments.

During my occupational therapy education, therefore, I reviewed the literature, and I came across the term *mild disabilities* which was used to describe children with attention deficits, developmental coordination disorder, and learning disabilities. I also came across the term *school-based occupational therapy*. I learned that the goal of occupational therapy in the school system is to improve the child's performance of tasks important for full participation in his or her student role. I also learned that for the past 35 years, the provision of both occupational and physical therapy in schools has been mandated by law in some world regions. In many of these regions, public schools have become the most common practice arena for occupational and physical therapists working with children. This, however, is not the case here in Sweden.

Because occupational therapists have not been employed by the community to work as related service providers here in Sweden, I became interested in how we might begin to make a change. Imagining a new practice arena for occupational therapists in Sweden, my wish became to gain a better understanding of school-based occupational therapy, the usefulness of assessments, and the effectiveness of interventions provided in the school setting with children without obvious physical, cognitive, or neurological disabilities. My passion became to advocate for our profession in relation to how we can contribute to the school health and educational teams when serving children with and without disabilities who have difficulties performing the tasks they want,

need, or are expected to perform during their school day to fully participate in their student roles.

Traditionally, occupational therapists have assessed children in therapy rooms or in clinical settings, but I believe that if we are to practice in the school context, we need to use instruments that are relevant for the educational setting, and provide services that focus on enhancing children's school-related task performance and participation at school. I also believe it is critical that we practice in a manner that stresses our unique focus on occupation by implementing occupation-based assessments and interventions rather than focusing on underlying performance components or specific diagnoses. My research questions focused, therefore, mainly on instrument development (validity and reliability) of a relatively new and relevant assessment tool used for measuring children's quality of schoolwork task performance in naturalistic classroom settings.

Every other day we can read in the newspaper about children with mild disabilities, and how they and their caregivers are discouraged or that children with mild disabilities underachieve at school. It is also well known that children with mild disabilities often "fall between the cracks," and do not receive needed services. I hope this thesis will broaden our knowledge and contribute to discussions about children with mild disabilities, and how we, together with other service providers, can address and support their occupational performance needs in school settings to prevent and minimize such long-term implications as dropping out of school and developing mental health problems in adolescence and adulthood.

Michaela Munkholm

Introduction

The primary focus of this thesis was the further development and validation of the School Version of the Assessment of Motor and Process Skills (School AMPS) (Fisher, et al., 2007), an observation-based evaluation of quality of occupational performance when children perform schoolwork tasks in school settings. My long term goal was to contribute to knowledge about children with mild disabilities who often experience difficulties with occupational performance in school settings, and how the School AMPS can be used when a true top-down process of planning and implementing school-based occupational therapy services is in a Swedish context.

Therefore, in the following sections I will first define occupation in relation to occupation performance, areas of occupation, and levels of doing. Then, I will discuss occupational roles and the relation between role expectations and the tasks people perform. Here, my emphasis will be on the student role. From there, I will progress to discussing both students with mild disabilities and those who have no diagnosed disability, but who experience problems in school settings. I will then proceed to a discussion of occupational-therapy-specific assessment and intervention services in school settings. Finally, because the focus of my thesis has to do with the further development and validation of the School AMPS, I will describe the School AMPS, and then briefly discuss Rasch measurement models, define the terms *reliability* and *validity*, and discuss how reliability and validity are evaluated from a traditional and a Rasch measurement perspective.

Occupation, doing, and occupational performance

In occupational therapy, the term *occupation* is a core concept of the profession. Occupation – the engagement in performance of daily life tasks that have meaning to the person – is believed to support or lead to participation in life (American Occupational Therapy Association [AOTA], 2008). Definitions of the term vary both within and outside the occupational therapy literature. In the Oxford

English Dictionary (1989) occupation is defined as *action* in relation to occupying a role, occupying time, or occupying space, or as being engaged in a particular pursuit or activity. “Occupation is everything people do to occupy themselves, including looking after themselves (self-care), enjoying life (leisure), and contributing to the social and economic fabric of their communities (productivity)” (Canadian Association of Occupational Therapists [CAOT], 1997). These definitions specify that something is happening – occupation pertains to someone *doing* something (Fisher, 2006, 2009). That something is the performance of any task that is both culturally and personally meaningful and/or purposeful for the person who is performing the task (cf., Fisher, 2009; Reed & Sanderson, 1999; Townsend & Polatajko, 2007; Wilcock, 1999). Thus, *occupational performance* pertains to engagement in a task performance that unfolds over time as a dynamic relationship between the person, the environment, and the task the person is performing (Kielhofner, 2008; Law, Baum, & Dunn, 2005; Law, et al., 1996; Townsend & Polatajko, 2007). When the person is a child, participating in the classroom requires the ability to perform tasks assigned by the teacher, and the child’s engagement in task performances becomes a dynamic relationship between the child, the classroom environment, and the task the child is performing.

Areas of occupation

The tasks of daily life that people perform can be related to different *areas of occupation* – groupings that seem to be based on the nature or characteristics of the tasks. For example, in the Occupational Therapy Practice Framework (AOTA, 2008), areas of occupation are defined as activities of daily living (ADL), instrumental activities of daily living (IADL), rest and sleep, education, work, play, leisure, and social participation. Elsewhere, daily life tasks are classified as self-care, productivity, and leisure (Townsend & Polatajko, 2007; CAOT, 1997). People engage in task performances associated with these different areas of occupations in a range of environments, including their homes, work places, schools, or communities.

In school settings, children engage in task performances related to self-care (e.g., putting on and taking off clothing when transitioning between the outdoors and the classroom, eating, toileting), work (e.g., doing chores within the classroom), play, social participation, and education (Bundy, 1995; Swinth, Chandler, Hanft, Jackson, & Shepherd, 2003). The area of occupation in focus in this thesis is

education, which refers to the “activities needed for being a student and participating in the learning environment” (AOTA, 2008). Task performances related to education include writing, cutting, coloring, pasting, doing mathematics – task performances that support learning.

Levels of doing

As I described above, occupation pertains to someone doing something – participation in life situations and engagement in task performances. As the doing unfolds, one can observe a chain of action (AOTA, 2002; Fisher, 2006, 2009; Kielhofner, 2008). Thus, occupation can be described as a continuum of levels of doing. Kielhofner (2008) referred to these levels as *occupational participation*, *occupational performance*, and *occupational skills*.

Occupational participation

The term occupational participation can be seen as the broadest level of occupation as it relates to the engagement or involvement in a variety of daily life tasks that are part of a person’s social and cultural context (AOTA, 2008; World Health Organization [WHO], 2001). It is also assumed that engagement in occupation provides structure to everyday life and contributes to health and well-being (Kielhofner, 2008; Townsend & Polatajko, 2007; Wilcock, 1999). Occupational participation is both personal and contextual in that a person’s engagement will not only depend on the individual’s strengths and weaknesses, motives for engagement, roles, and habits (Kielhofner, 2008), but also on the resources and demands in the environment that support or constrain participation in occupation (Townsend & Polatajko, 2007). For a child within the school setting, occupational participation pertains to his or her ability to be involved and engage in the school tasks performances needed to fulfill his or her student role. Thus, throughout the remainder of this thesis, I will use the term *student* to refer to a child when he or she is performing tasks within the school setting.

Occupational performance

Occupational performance refers to the actual *doing* or performance of specific daily life tasks (e.g., writing a sentence, eating a meal, playing soccer) (Kielhofner, 2008). The performance of a task is shaped by culture, local customs, and personal habits which become

incorporated into one's ways of doing. The requirement for and specific form of occupational performance can also be established by law (e.g., performance of tasks in relation to mandatory education). The environment also impacts a person's occupational performance not only in the sense of support and constraints provided by the objects used when performing a task, but also in terms of the spaces within which the task is performed. For a child in a school setting, factors that influence occupational performance include the culture of the classroom, implicit and explicit rules established by the teacher or the educational system, the specific requirements of the task the teacher has assigned, where tools and materials are located, which tools and materials they are to use, and so on. For example, a specific occupational performance within the classroom might pertain to writing some sentences related to a book the students in the class have just read. The teacher may further specify that the students are to use pencils to write four sentences, and when they are done, put their finished writing on the teacher's desk and any tools and materials they have used (e.g., pencils, erasers) back in their desks.

Occupational skills

Occupational performance – the performance of a task – is comprised of meaningful *sequences of actions* that unfold over time. That is, when a person performs a task, he or she performs a chain of actions that are called occupational or performance skills. These performance skills comprise of the smallest observable units of the *doing* that emerge in the process of performing a task (Fisher, 2006, 2009; Fisher & Jones, 2010; Kielhofner, 2008). For example, as a student is performing a writing task, we can observe him or her *reaching* for a pencil, *grasping* the pencil, *repositioning* the pencil in her hand, and *starting* to write. As he or she is *writing*, we might observe him or her *pausing*, *looking* at the teacher, and *asking* the teacher a question, “How many sentences did you ask us to write?” Then, after the teacher has answered, we can see him or her *continuing* to write, and *stopping* after he or she has written four sentences.

The concept of *performance skills* originated with the development of the Assessment of Motor and Process Skills (Fisher, 2006; Fisher & Jones, 2010) and later they were incorporated into the Model of Human Occupation (Kielhofner, 2008), the Occupational Therapy Intervention Practice Model (OTIPM) (Fisher, 2009), and the Practice

Framework (AOTA, 2002). The *performance skills* are divided into motor skills, process skills, and social interaction skills. In the example above, *reaching* for, *grasping*, and *repositioning* the pencil in the hand are motor skills. *Starting* to write, *pausing* and then *continuing* to write, and *stopping* are process skills. *Looking* at the teacher and *asking* a question are social interaction skills.

Childhood roles and the interaction between occupation and roles expectations

As I noted above, when a child is participating in the classroom, performing tasks assigned by the teacher, he or she is in his or her student role. All daily life task performances are closely related to desired or expected roles, and each role has a defined status and specific expectations for desired behaviors (Rodger & Ziviani, 2006). Roles are dynamic and change over time, both during the course of the day and with development or skill acquisition. As the person develops and acquires new skills, new roles emerge and others are replaced over one's lifespan (Reed & Sanderson, 1999).

For example, while play is considered the primary occupation of young children, their primary role as player changes when the child begins school (Case-Smith, Law, Missiuna, Pollock, & Stewart, 2010; Rodger & Ziviani, 2006). In fact, because children are required to spend considerable time at school, their primary role transitions to that of student (Rodger, 2010). While roles shape what children do and influence their identities, there is also a transaction between the role expectations of the students and the tasks they must perform. That is, when children are at school and are engaged in their student roles, they must effectively perform schoolwork tasks as well as a variety of other school-related nonacademic daily life tasks, including ADL, play, and social participation (Bundy, 1995; Chen & Cohn, 2003; Rodger, 2010). Moreover, difficulties in performing daily life tasks due to a disability or impairment may contribute to difficulties in role performance, and as a result, a conflict between one's own or others' role expectations may occur (Kielhofner, 2008). Thus, a child is able to act and fully participate in the student role only to that extent that he or she can perform effectively those tasks integral to defining that role.

The studies in this thesis originally emerged from (1) my observations of students who struggled in their student roles and/or demonstrated difficulties in performing daily life tasks in classroom settings, and (2) the questions I had related to the lack of support for children with and without diagnosed mild disabilities. Therefore, in the next section, I will proceed to define and describe students with mild disabilities as well as those who have no diagnosed disability, but who experience problems in school settings. In this thesis, I will refer to these latter students as students who are at risk. I will then discuss occupational-therapy-specific assessment and intervention services in school settings in relation to the needs of students at risk or with mild disabilities.

Students with mild disabilities or who are at risk

Students with mild disabilities are defined in this thesis as those classified by professionals as having disorders of attention (e.g., attention-deficit/hyperactivity disorder [ADHD]), specific learning disabilities (LD) (e.g., reading disabilities, speech and language impairments), developmental coordination disorder (DCD) (American Psychiatric Association [APA], 2004), or sensory integrative dysfunction (SI) (Bundy, Lane, & Murray, 2002). The classification of ADHD, LD, and DCD is based on symptoms or patterns of behaviors that (1) need to be present in different settings to a degree which is inconsistent with the child's current developmental level; (2) are not explained by other physical or cognitive impairments or medical diagnoses; and (3) interfere with the child's school, social, and occupational functioning (APA, 2004). SI is a disorder defined within sensory integration theory and practice, an approach often used by occupational therapists to "diagnose" or describe children who have difficulties processing sensory information and planning and organizing movement. Very often, these children are described as being "clumsy." They may also display learning disabilities and other behavioral problems such as distractibility (Bundy, et al., 2002). Thus, their symptoms are often very similar to those who have DCD. Moreover, DCD often coexists with ADHD and LD (APA, 2004; Dewey, Kaplan, Crawford, & Wilson, 2002; Dewey & Wilson, 2001), and evidence suggests that overlap between conditions is more common than having one condition alone (Kaplan, Wilson, Dewey, & Crawford, 1998; Kopp, 2010).

Students with mild disabilities are often overlooked when it comes to design and implementation of adequate supports in school settings (Cook, 2001; Skolverket, 2001a, 2001b). This, in part, may be because it is common to prioritize supports for students with more severe, visible physical disabilities. More importantly, from an occupational therapy perspective, is the fact that even though mild disabilities are not always visibly apparent, and despite their subtle and/or obscure nature, they commonly have a negative impact on the students' occupational performances (Cunningham & Boyle, 2002; Fox & Lent, 1996; Missiuna, Rivard, & Pollock, 2004). That is, even though they may have different diagnostic labels, their symptoms are similar when it comes to occupational functioning in the school environment. More specifically, children with mild disabilities often struggle to perform even the simplest tasks related to maintaining their student role (e.g., attend to class, follow rules, organize their workspace, use objects effectively, complete assignments on time, socialize with peers) (Missiuna, et al., 2004). This negative impact on occupational performance typically persists throughout adolescence and adulthood (Cantell, Smyth, & Ahonen, 1994; Hellgren, Gillberg, Gillberg, & Enerskog, 1993; Losse, et al., 1991). Additionally, it is common that students with mild disabilities experience failure, come to expect failure, and develop lower self esteem that affects their social, academic, and physical performances (Rasmussen & Gillberg, 2000; Schoemaker & Kalverboer, 1994; Skinner & Piek, 2001). Thus, long term implications such as dropping out of school and developing mental health problems in adolescence and adulthood are evident (Kadesjö & Gillberg, 1998; Rasmussen & Gillberg, 2000).

As I noted above, students at risk are defined in this thesis as students who experience problems in the classroom, but who have not been identified as having a disability (Heward, 2000). Such students have been described as being "virtually identical" behaviorally and cognitively to students with mild disabilities (Johnson, 1998), as they are typically characterized as having problems with memory and cognition, language, attention, social skills, academic performance, and behavior.

Like students with mild disabilities who are often overlooked and "fall between the cracks," (Socialstyrelsen, 2003) there also is a high probability that students at risk fail to receive needed supports in school settings. In fact, it is often the case that students at risk are not eligible for occupational therapy or other related services until it

is determined that they have a diagnosis and/or test scores that are below age expectations. Occupational therapists, however, do sometimes evaluate such students in order to determine if schoolwork performance, developmental milestones, gross or fine motor coordination etc. are below age expectations (Case-Smith & O'Brien, 2010).

Reasons for referral to occupational therapy

One of the most common reasons students at risk or with mild disabilities are referred to occupational therapy for evaluation and/or intervention is because of difficulties with handwriting (Schneck & Amundson, 2010). Other common teacher concerns for referral to occupational therapy services are related to attention, organizational difficulties, and/or clumsiness (e.g., using an odd pencil grip, writing takes a lot of effort, difficulty following instructions, getting started and finishing tasks on time, dropping things or bumping in to obstacles). Teachers also often report concerns related to behavior in the classroom, such as coping with classroom rules or difficulties interacting socially with peers (Dewey, et al., 2002; Dussart, 1994; Missiuna, et al., 2004).

Occupational therapy assessment and interventions in school settings

When occupational therapists work in the school settings, the focus becomes the child's ability to perform school tasks required to be able to participate in his or her student role (Rodger, 2010). In fact, legislation about occupational therapy as a related service in the United States mandates that the focus of all occupational therapy services provided in the school setting must be educationally relevant and related to the student role (Individuals with Disabilities Education Act [IDEA], 2004).

Occupational therapy as a related service

Not only must occupational therapy services provided in the school setting be related to the student role, occupational therapists in some countries are mandated to provide services in the school setting as part of the national or local educational policies (e.g., education acts) such as the Individuals with Disabilities Education

Act (IDEA) (2004). In other countries, including Sweden, there are also national policies that urge, but do not mandate, that school-based external support services (e.g., occupational therapy, physical therapy) be available for students with special educational needs (UNESCO, 1994). Apart from how services are legislated, or whether or not occupational therapy services are integrated into the educational system, there are some important commonalities regarding the rights of students with disabilities. These rights assert that students with disabilities are to be included in regular education classrooms, education must be for all children, and supports for students with special educational needs should be provided irrespective of medical diagnosis or other impairments (IDEA, 2004; Svensk Författningssamling [SFS], 1995).

Ecologically relevant and occupation-based occupational therapy services

Because school-based occupational therapy services are to be related to the student role, this impacts on the evaluation and intervention methods that are appropriate for use in school settings. One fundamental feature that has to be considered is the ability to evaluate the student and intervene within the natural school context. This requires using an ecological approach, taking into consideration the interaction between the child and the environment during participation in occupation – performance of school tasks (Kramer & Hinojosa, 2010; Law, et al., 1996). The environment includes physical, social, cultural, and temporal aspects that impact on occupational performance and choice of occupations (Dunn, Brown, & McGuigan, 1994; Fisher, 2009; Kielhofner, 2008).

Following an ecological approach for evaluation and intervention in the natural school environment means that services are implemented in the classroom, during the student's typical routines. Because the occupational therapy evaluations are based on observing and evaluating the student's quality of occupational performance, the evaluation process becomes occupation-based. Moreover, the process of evaluating, and then planning and implementing interventions, are made in collaboration with others (e.g., parents, teachers, other service providers) (Dunn, 2000; Ziviani & Rodger, 2006). By focusing on the student role, evaluating the quality of a student's occupational performance within the natural educational setting,

and then developing relevant intervention plans, occupational therapists can contribute to students' developing and improving their quality of educational and functional school task performance (Bundy, 1995; Case-Smith & O'Brien, 2010; Swinth, et al., 2003).

Top-down versus bottom-up approach to assessment

The preferred approach to assessment in the school setting is to use of a top-down and occupation-based approach (Brown & Chien, 2010; Case-Smith & O'Brien, 2010). This approach focuses first using client interviews to gather information about the student's occupational roles, reported quality of occupational performance, and the student's level of engagement and participation in the natural school context (Case-Smith & O'Brien, 2010). During the initial interview with the student's teacher, the occupational therapist may also choose to use the School Function Assessment (SFA) (Coster, Deeney, Haltiwanger, & Haley, 1998) to gather information about the student's level of participation as reported by the teacher.

Once the occupational therapist has determined what occupational performances are strengths for the student and which ones the teacher is concerned about and present challenges for the student, the next step in a top-down approach is to use informal observation or a standardized observational assessment to evaluate the student's quality of occupational performance. If the types of tasks that the teacher has prioritized are schoolwork tasks, the occupational therapist may choose to administer the School AMPS, a standardized, ecologically-relevant, observation-based assessment of a student's quality of schoolwork task performance in natural classroom settings (Fisher, 2009; Fisher, et al., 2007). I will describe the School AMPS in more detail below.

In summary, a top-down approach to assessment begins with obtaining a broad picture of the client (e.g., the student and his or her teacher), the context of the student's occupational performances (including the nature of the classroom in which they occur), and the task performances that the client considers to be of most concern; and then progresses to an observation of the student's quality of occupational performance in the natural classroom setting (Fisher, 2009; Fisher, et al., 2007). If, after implementing ecologically-relevant and occupation-based assessments, the occupational

therapist determines that there is a need to further clarify what is the underlying problem – the cause of the student’s diminished quality of schoolwork task performance – the occupational therapist may choose to evaluate underlying impairments or body functions (WHO, 2001). Such assessment tools, however, should only be used when the cause of the occupational performance problem is not apparent and if there is a need for further clarification (Dunbar, 2007; Hinder & Ashburner, 2010; Hocking, 2001; Law, et al., 2005).

This is in contrast to traditional occupational therapy services which have been based on a bottom-up, medical model (Brown & Chien, 2010; Hinder & Ashburner, 2010). When a bottom-up approach is used, the evaluation is focused on identifying the student’s underlying impairments (e.g., poor fine motor coordination, distractibility) (Fisher, 2009). In fact, many occupational therapists working in school settings continue to use a traditional, direct service, pull-out delivery model for both assessment and intervention (Bayona, McDougall, Tucker, Nichols, & Mandich, 2006; Case-Smith & Cable, 1996; Polatajko & Cantin, 2010; Rodger, Brown, & Brown, 2005; Spencer, Turkett, Vaughan, & Koenig, 2006). When this occurs, the focus of the occupational therapy evaluation often becomes the student’s impairments of underlying body functions, not occupational performance (Mandich, Polatajko, Macnab, & Miller, 2001; Rodger, et al., 2005).

In fact, majority of the standardized assessment tools that are most commonly used by occupational therapists working in school settings remain those designed to evaluate body functions (e.g., motor coordination, visual perception, sensory processing) despite the fact that the results of such assessments do not clarify the student’s problems with occupational performance – the consequence of the interaction between the student, the task, and the environment. That is, such assessment tools do not relate directly to the student role nor the student’s performance of those tasks that support his or her student role and enable participation (Burtner, McMain, & Crowe, 2002; Coster, 1998; Sangster, Beninger, Polatajko, & Mandich, 2005; Weintraub & Kovoshi, 2004). This unfortunate situation may have occurred because there are very few assessment tools available that are ecologically-relevant, occupation-based, and designed to be used to measure the student’s occupational performance issues in the school context. One of the few available tools is the School AMPS (Fisher, et al., 2007). As the further development and validation of the School AMPS was the

overall aim of this thesis, I will describe the School AMPS in more detail in the section that follows.

School Version of the Assessment of Motor and Process Skills (School AMPS)

As I mentioned above, the School AMPS (Fisher, et al., 2007) was designed to be used by occupational therapists to measure the student's quality of performance of common schoolwork tasks (e.g. cutting with scissors, coloring, using the computer) in naturalistic classroom settings. Thus, the School AMPS is a standardized, occupation-based, and ecologically-relevant observational evaluation tool. The School AMPS is also client-centered in that the student is always observed and evaluated when performing schoolwork tasks that the client (e.g., teacher and student) has identified as presenting a challenge for the student. By using the School AMPS for evaluation, the occupational therapist is able to (1) determine if the student needs occupational therapy services, (2) establish occupation-based goals, (3) plan occupation-focused interventions to enhance the student's occupational performance in the classroom, and (4) perform follow-up evaluations to determine if the provided occupational therapy services were effective in improving the student's schoolwork task performance.

The School AMPS evaluation is carried out within the students' regular classroom environment, during normal classroom routines, and the School AMPS can be used to test all students, 3 years of age and older, regardless of diagnosis, impairments, or other reasons for difficulties with performing schoolwork tasks. In the School AMPS, there are 25 schoolwork tasks that range in challenge. These tasks reflect the most commonly performed schoolwork tasks in pre- and primary school settings. These schoolwork tasks have been grouped into five categories: Drawing & coloring tasks, Pencil/pen writing tasks, Cutting & pasting tasks, Computer writing tasks, and Math manipulative tasks. A unique feature of the School AMPS is that the evaluation focus is on the student's quality of schoolwork task performance, not his or her underlying impairments or body functions. Additionally, the items in the School AMPS represents the smallest observable units of the *doing* – the school motor and school process performance skills that are linked together one by one as the student enacts the schoolwork task.

When implementing a School AMPS evaluation, the occupational therapist follows a top-down approach to assessment and interventions that is based on the Occupational Therapy Intervention and Process Model (OTIPM) (Fisher, 2009). The OTIPM provides a structure to guide professional reasoning to ensure that the occupational therapist is using a top-down, client-centered, and occupation-based approach to assessment and intervention. The procedure of being client-centered provides a foundation for establishing a collaborative relationship with the teacher, student, parents, and sometimes other persons working with the student (e.g., special education teacher, members of the school health team, after school program teachers). Such collaborative practice among the educational team members is necessary for successful inclusive education and effective occupational therapy in schools (Barnes & Turner, 2001; Bose & Hinojosa, 2008; Bundy, 1995).

The occupational therapist initially begins to establish a collaborative relationship with the teacher in the context of interviewing the teacher and learning about the student and his or her strengths and challenges in the classroom, the teacher's concerns, and the overall context of the student's schoolwork task performance. This phase is the first step described in the OTIPM (Fisher, 2009), and thus, is also the first step in the administration of the School AMPS evaluation (Fisher, et al., 2007). Part of establishing the context of the student's schoolwork task performance is to gain an understanding about the student's school performance from the perspective of the teacher, as well as to learn about the physical and social performance context of the school (e.g., available societal resources such as special education services or aids in the classroom, students' social relationships, daily routines, classroom rules, teacher expectations).

Among the schoolwork tasks that the teacher has identified as challenging for the student to perform, the teacher then prioritizes which schoolwork tasks are most appropriate for the occupational therapist to observe. The occupational therapist also learns about other specific criteria of the tasks assigned to the students (e.g., if the students are expected to use pencils, crayons, or markers to write; if the students are allowed to share tools and materials; how do the students call for attention when they need help; what are the students to do with their finished work when they are done).

After the interview, the occupational therapist schedules a time to observe the student when he or she performs at least two School AMPS tasks that the teacher has identified as presenting a challenge for the student. The observation always takes place within the natural classroom setting without disrupting the normal classroom routines, and the classroom can be any room in which the student typically performs schoolwork tasks provided there are at least three other students and a teacher present.

After the observation, the occupational therapist scores the quality of the students' performance on 16 school motor skill items and 20 school process skill items for each School AMPS task performed. These skill items are motor and process performance skills – the goal-directed actions that represent the smallest observable units of occupational performance that are linked together as the student performs the schoolwork task (e.g., writing four sentences) (Fisher, 2006, 2009; Kielhofner, 2008).

The scoring of the School AMPS is criterion-referenced and each skill item (action) is scored on a 4-point rating scale according to the standardized criteria outlined in the School AMPS manual (Fisher, et al., 2007). The student is given lower scores if he or she demonstrates increased clumsiness or effort, decreased efficiency, decreased safety, and/or need for assistance during the performance of the individual goal-directed actions when completing each schoolwork task. Additionally, the classroom environment and the overall quality of schoolwork task performance are rated for each schoolwork task performed.

After scoring, the occupational therapist enters the school motor and school process skill item raw scores, for each schoolwork task observed, into the School AMPS computer-scoring program (Three Star Press, 2005) in order to generate the student's school motor and school process quality of schoolwork performance measures. The occupational therapist then uses the information gathered during the interview with the teacher and the results of the School AMPS observation to collaboratively, together with the teacher, write client-centered and occupation-based goals, and to plan and implement occupation-based interventions. Finally, once the planned interventions have been implemented, the School AMPS can again be used to evaluate if change has occurred.

Development of the School AMPS – issues of reliability and validity

Before initiating the evaluation process, there are several issues that an occupational therapist has to consider. That is, when choosing and using a test or instrument, it is crucial to determine for which purpose the instrument is to be used, and if there is evidence that the test scores or measures are reliable and valid for that purpose (Cook & Beckman, 2006). Additionally, if one purpose is to measure change, sensitivity and reliability of the test scores or measures also must be considered. If our outcome measures are not sensitive and reliable, then we cannot accurately measure change over time. Moreover, reliability and validity of an instrument are also essential for making correct interpretations of the measurement results (Downing, 2003; Law, Baum, & Dunn, 2001; Law, et al., 2005).

As described earlier, the School AMPS quality of schoolwork performance measures are generated using the School AMPS computer-scoring program (Three Star Press, 2005) which is used to implement a many-facet Rasch (MFR) analysis (Linacre, 1993) of the student's school motor and school process skills item raw scores. As the four studies that comprise my thesis relate to evaluating aspects of reliability and validity, I will now proceed to introduce briefly Rasch measurement methods and the MFR model that was used to develop the School AMPS. I will then proceed to define the terms *reliability* and *validity*, and then I will discuss how they are evaluated from a traditional and a Rasch measurement perspective. I will describe MFR analyses of School AMPS data in more detail in the methods section.

Development of instruments using Rasch models

The family of Rasch models offers an approach to instrument development and validation (Bond & Fox, 2007). Rasch models are based on assertions that are probabilistic in nature (Bond & Fox, 2007; Wright & Masters, 1982). More specifically, a person's performance on an item can never be predicted with 100% accuracy and Rasch models acknowledge the probabilistic nature of human behavior (Wright & Linacre, 1989). However, if the data meet the requirements of the respective Rasch model (e.g., data fit the model), each person's ability measure and each item's difficulty calibration can be positioned on a common unidimensional scale of measurement. Thus, items are positioned along the linear

measurement continuum according to their difficulty (e.g., arranged from those that are easy to perform to those that are hard to perform), and the persons are positioned on the same continuum according to their ability to perform those items (Bond & Fox, 2007). Additionally, in the development of a test based on Rasch analysis methods, items that do not follow the basic assertions of the respective Rasch model must be considered for omission from the test. Finally, computer programs implementing Rasch measurement models generate statistics that can be used to test how well the data fit a chosen model as well as aspects of reliability of the estimated measures.

The many-facet Rasch model for the School AMPS

As I noted above, the Rasch model for the School AMPS is a many-facet model (Fisher, 1993; Linacre, 1993; Linacre & Wright, 2002). More specifically, there are four facets – student ability, rater severity, task challenge, and item difficulty. When students' ordinal school motor and school process skill item raw scores are subjected to MFR analyses so as to be converted into linear (equal-interval) quality of schoolwork performance measures expressed in logits (log-odds probability units), they are adjusted to account for the severity of the occupational therapist who rated the student's quality of performance, the challenges of the schoolwork tasks the student performed, and the difficulty of the School AMPS items (Fisher, et al., 2007). There are two School AMPS scales, one comprised of the School AMPS motor skill items (school motor scale) and one comprised of the School AMPS process skill items (school process scale). The MFR logistic transformation of students' skill item raw scores makes it possible to position each student along two linear continua of "ability," one for school motor and one for school process. Thus, higher quality of schoolwork performance measures represent more able students and lower quality of schoolwork performance measures represent less able students. Likewise, skill items and schoolwork tasks with higher logit values are easier to perform, and skill items and schoolwork tasks with lower logit values are harder to perform.

Aspects of reliability

Reliability refers to stability or consistency of measurement when a testing procedure is repeated on an individual or on the same group of individuals (American Educational Research Association,

American Psychological Association, & National Council on Measurement in Education, 1999). Some variation is always present between repeated testing, and the extent of this variation is referred to as error of measurement (Crocker & Algina, 1986). If the measurement error is too large, then real changes cannot be distinguished from measurement error. Sources of measurement error can be categorized as components related to individual factors (e.g., motivation, anxiety), characteristics of the measure related to the instrument (e.g., items or tasks), or the circumstances under which the instrument is used (e.g., administration, rater scoring, unclear instructions, time of day) (Crocker & Algina, 1986; Haertel, 2006). The most common ways to report reliability are as reliability coefficients and as standard errors of measurement (the standard deviation of measurement errors [*SEm*]) (Haertel, 2006). While reliability coefficients are used to compare consistency of measures from different evaluation procedures (e.g., different occasions, different forms of the instrument), the *SEm* provides an estimate of how much an individual's measures are expected to vary around his or her true score (American Educational Research Association, et al., 1999; Crocker & Algina, 1986). In general, a high reliability coefficient and a low *SEm* indicate that scores are stable or consistent.

There are several different procedures for estimating the degree of inconsistency caused by random errors (e.g., test-retest, alternate forms, equivalent or parallel forms, split-half reliability methods) (Allen & Yen, 2002). The test-retest method is used to establish a coefficient of stability and therefore requires that two or more test administrations are performed within a reasonable time frame (Cohen & Swerdlik, 2010; Crocker & Algina, 1986). One disadvantage of using this method is the risk for carry-over effects between evaluation times (Allen & Yen, 2002). Therefore, equivalent or parallel forms reliability is recommended as the best method to evaluate reliability, especially if two different forms of the same instrument are administered to the same individuals at different times (Ary, Jacobs, & Razavieh, 2002). Thus, two parallel forms of the same instrument are used to evaluate the coefficient of equivalence, and if administered at different times, the coefficient of stability and equivalence (Cronbach, 1947). However, the disadvantage of the parallel forms method is that this method has certain requirements for the two forms to be strictly parallel (e.g., equal means, equal variances) (Feldt & Brennan, 1989). In situations where only one test administration is possible, one can use the split-

half methods to estimate reliability (Allen & Yen, 2002; Crocker & Algina, 1986). Traditionally, when split-half methods are used, the items in a test are divided into two halves with the intention of creating two shorter forms that are parallel or as parallel as possible (Becker, 2000; Haertel, 2006). Thus, the correlation of equivalence is established, but now based on two halves of the same instrument (Ary, et al., 2002; Crocker & Algina, 1986; Cronbach, 1951; Haertel, 2006; Schmidt, Le, & Ilies, 2003). If the two halves are parallel (equivalent), the Spearman–Brown formula can be used to calculate the reliability of the total length instrument (Allen & Yen, 2002; Cohen & Swerdlik, 2010).

As discussed earlier, Rasch analysis computer programs generate statistics that can be used to test how well the data fit a chosen model. In this process, reliability estimates are generated for both the items and the persons in the form of the Rasch equivalent of Cronbach’s alpha (Wright & Stone, 1999) and *SE*ms. The Rasch equivalent of Cronbach’s alpha is based on a single administration of a test, and when using Rasch measurement methods, the *SE*m is estimated for each person ability measure and each item difficulty (Linacre, 2007; Schumacker & Smith, 2010; Wright & Masters, 1982; Wright & Stone, 1979). When using the estimated measures to evaluate change, these *SE*ms can be used to describe the confidence intervals within which each person’s true ability falls (Schumacker & Smith, 2010; Wright & Masters, 1982).

Aspects of validity

In the process of developing and evaluating tests or instruments, the most fundamental consideration is validity (American Educational Research Association, et al., 1999). The conceptualization of validity has emerged over the past 70 years, and has gone from being categorized into three specific types: content validity, criterion-related validity, and construct validity, to being defined as “the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests” (American Educational Research Association, et al., 1999, p. 9). Thus, in the validation process, the test developer or test user must collect and evaluate different aspects of validity evidence in relation to the soundness of the instrument and the measures or total scores that are obtained when the instrument is used to evaluate persons (i.e., test content, response processes, internal structures, relations to other variables, and consequences of testing) (Cook & Beckman, 2006; Downing,

2003). In the studies that comprise this thesis, we have collected evidence related to *internal structure* (Studies II and III), *relations to other variables* (Studies III and IV), and *consequences of testing* when using the School AMPS (Fisher, et al., 2007) in clinical practice or for research purposes (see Table 1).

Differential item functioning

One aspect of scale validity is to ensure that an instrument is free of differential item functioning (DIF). DIF occurs when groups (e.g., students in different world regions), who are supposed to be comparable on the construct being measured, display significant differences in their responses to a particular item (American Educational Research Association, et al., 1999; Tennant & Pallant, 2007). When using Rasch measurement methods, we would expect that the items retain their order of difficulty irrespective of which person is evaluated using the test. More specifically, as illustrated in Figure 1a, the school motor skill items are equally difficult to perform for students in Region A as for students in Region B. Thus, we can say that the hierarchy of the skill item difficulties is stable between regions. This can be compared to Figure 1b, where *Lifts* is harder to perform for students in Region A than for students in Region B. Thus, Lifts displays DIF between these two regions.

Table 1. Validity evidence collected in Studies II to IV

Validity evidence	Study II	Study III	Study IV
Based on internal structure	Evaluation of stability of skill item hierarchies (DIF)	Evaluation of stability of skill item hierarchies (DIF)	
Based on relations to other variables		Evaluation of sensitivity to detect change between groups expected to differ	Evaluation of responsiveness to change following intervention (sensitivity)
Based on consequences of testing	Evaluation for test bias related to test fairness (DTF)	Evaluation of student strengths and limitations to better target interventions	

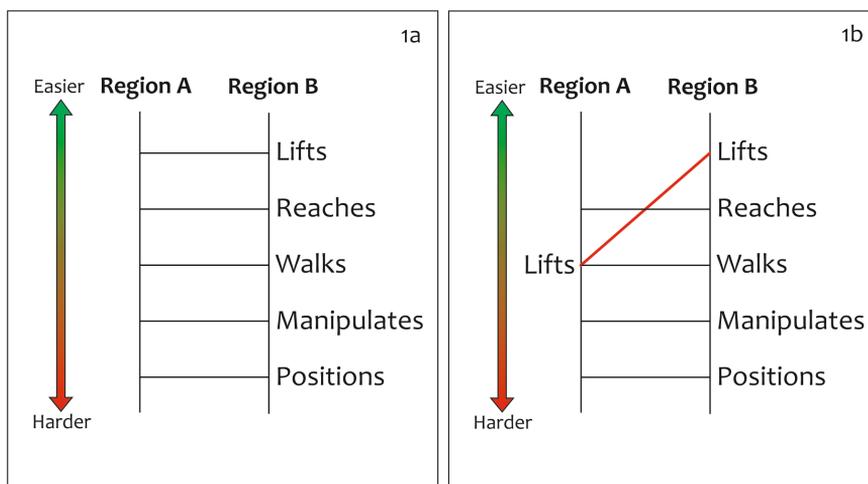


Figure 1a and 1b. Relative item difficulty calibrations for Region A versus Region B. In Figure 1a, the item difficulty hierarchy remains stable, but in Figure 1b, the item difficulty calibration for Lifts differs between regions.

From a validity perspective, DIF is a concern because it suggests that students from one group (e.g., world region) have differing probabilities of success on an item, placing one group at an advantage over another (Crocker & Algina, 1986; Osterlind & Everson, 2009). DIF analyses, however, also must include investigation of whether individual items function differently for different groups due to construct-irrelevant factors (Camilli, 2006; Pennfield & Camilli, 2007; Perrone, 2006) (e.g., if *Lifts* is found to be due to a factor that is irrelevant to the construct the School AMPS is intended to measure – quality of schoolwork task performance). To use a rather extreme example, if any skill item in the School AMPS displaying DIF were also shown to reflect comprehension of the Swedish language, then students from English-speaking countries would be at an unfair disadvantage as the item captures more than just quality of schoolwork task performance, it also captures knowledge of the Swedish language. In such a case, we would have evidence that DIF resulted in of lack of fairness of a *test item* (Camilli, 2006; Pennfield & Camilli, 2007).

Differential test functioning

In this thesis, differential test functioning (DTF) refers to equality of testing outcomes. Thus, a test may be considered to be “fair” if a total test score predicts the same level of performance for individuals of all groups (American Educational Research Association, et al., 1999). Said in other terms, while DIF alone does not always result in a risk to *test fairness* (Camilli, 2006), DIF becomes a special concern when it leads to unfair comparisons between groups based on total test scores or measures (Boomsma, van Duijn, & Snijders, 2001; Osterlind & Everson, 2009; Pae & Park, 2006; Wright & Stone, 1999). Moreover, according to Tennant and Pallant (2007), “irrespective of the amount of DIF detected, we argue that for practical purposes, given satisfactory fit to the model, if the person estimates remain largely unchanged, then the DIF is trivial and can be ignored. We define a trivial impact as being a difference in the person estimates from the two analyzes of less than 0.5 logits” (p. 1082). When the person estimates remain essentially stable, a test can be said to be free of DTF due to DIF (i.e., there is no risk to *test fairness* associated with DIF). In other words, the School AMPS would be valid for use irrespective of in which world region the student lives.

Sensitivity of measures

Another important aspect of validity pertains to evidence that the test scores or measures are sensitive enough to be used to identify differences between groups that are expected to differ on the trait being tested (e.g., students with and without mild disabilities). Such evidence can provide support for the use of the School AMPS when used to identify those students who might be in need of services. When sensitivity is considered from the perspective of providing occupational therapy services to children in schools, it also becomes important to determine if the School AMPS measures are sensitive enough to detect changes following occupational therapy intervention (American Educational Research Association, et al., 1999; Unsworth, 2000).

Validity can also pertain to consequences of testing. One such potential consequence that was considered in this thesis pertained to determining if the identification of similarities and differences in the skill profiles of typically-developing children versus those with mild disabilities might provide evidence that could inform practice. That

is, we speculated that if we could identify which performance skills (i.e., School AMPS skill items) were strengths (i.e., those that were essentially “intact” – the two groups have equal performance skill levels), and those performance skills that were “deficit” (i.e., those where the students with mild disabilities demonstrate less skill than do their typically-developing peers), we might be able to gather evidence that can inform occupational therapists about how to better plan their interventions.

Previous evidence for reliability and validity of the School AMPS

The School AMPS has been standardized internationally, and is currently being used in North America (NA), Australia and New Zealand (ANZ), United Kingdom and the Republic of Ireland (UK), Nordic countries, other areas of Europe (e.g., the Netherlands, middle Europe), and to a lesser extent in Central and South America and Caribbean (CSA) and Asia. Previous research has shown overall high inter- and intra-rater reliability as demonstrated by goodness of fit of the raters to the many-facet Rasch (MFR) model of the School AMPS ($MnSq \leq 1.4$ and $z < 2$) (Fisher, et al., 2007). Several studies have provided validity evidence based on internal structure of the School AMPS scales (skill items and tasks) (Atchison, Fisher, & Bryze, 1998; Fisher, Bryze, & Atchison, 2000; Fisher, et al., 2007).

The School AMPS tasks and all of the School AMPS skill items except Walks, Endures, and Paces have been shown to be free of differential task and item functioning, respectively, among three world regions (NA, ANZ, and Europe – UK, Nordic countries, and other Europe combined). Moreover, what DIF was identified did not result in DTF. The School AMPS tasks and skill items also have been shown to be free of differential task or item functioning between genders. Finally, there is evidence that quality of schoolwork task performance increases with age (Fisher, et al., 2007).

Previous validity evidence based on relations to other variables has shown that the School AMPS quality of schoolwork performance measures are sensitive enough to differentiate among groups of students who are expected to differ – (1) among typically-developing students, students who have identified developmental/neurological, cognitive/psychological, or other/multiple disabilities (Fisher, et al., 2007); and (2) between typically-developing students and students at

risk (Fisher & Duran, 2004). In addition, recent research has supported no gender differences in quality of schoolwork task performance among typically-developing students or among students at risk or with mild disabilities (Sperens, Munkholm, & Fisher, 2010).

When the existing evidence is considered in relation to the need for further research, the following conclusions can be drawn: While there is evidence to support rater reliability, there is no reported evidence related to the reliability of the School AMPS quality of performance measures. Moreover, if the School AMPS is to be used in Sweden or other Nordic countries there remains a need to ensure that the School AMPS skill items are free of cross-regional DIF, and if cross-regional DIF is detected, that it does not result in DTF when students living in the Nordic countries are evaluated. Finally, if the School AMPS is to be used to identify diminished quality of schoolwork task performance among students at risk or with mild disabilities or as an outcome measure following occupational therapy interventions provided in school settings, there is a need to generate evidence that the School AMPS measures are valid for such purposes.

Aims of this Thesis

The overall aim of this thesis was, therefore, to evaluate different aspects of reliability and validity of the School AMPS measures and scales in order to gather evidence related to the clinical usefulness of the School AMPS when evaluating students' quality of occupational performance in school settings.

Specific research aims

- To estimate the reliability of the School AMPS quality of schoolwork performance measures (Study I)
- To determine if the School AMPS is free of regional differential item functioning (DIF) when used to evaluate students from different world regions (Study II)
- To determine if the School AMPS is free of DIF when used to evaluate typically-developing students and students with mild disabilities (Study III)
- If there is evidence of DIF among world regions, to test if the School AMPS measures are free of differential test functioning (DTF) (Study II)
- To determine if students with mild disabilities actually demonstrate significantly greater challenge performing individual school motor and school process skills than do typically-developing students, and if so, which School AMPS skill items are actually performed less skillfully (Study III)
- To evaluate if the School AMPS measures are sensitive enough to identify differences in quality of schoolwork task performances between typically-developing students and students with mild disabilities (Study III)

- To evaluate if the School AMPS measures are sensitive enough to detect change in students' quality of schoolwork task performances after provision of occupational therapy interventions in school-settings (Study IV)

Methods

Participants

The participants in this thesis were students of varying ages and world regions who were (1) *typically-developing* students with no known diagnosis, (2) students *at risk* (i.e., students without identified diagnoses, but whose teachers had expressed concerns related to behavioral problem or risk for academic failure) (Heward, 2000; Johnson, 1998), and (3) students with *educational* (e.g., reading disorder), *medical* (e.g., autism, cerebral palsy), or *occupational therapy diagnoses*¹ (i.e., sensory integrative disorder) (Bundy, et al., 2002) (see Table 2). More specifically, in this thesis, students at risk were defined as students who were experiencing problems in the classroom, but who had not been identified as having a disability (Heward, 2000). As I described earlier in the introduction, such students have been described as being “virtually identical” behaviorally and cognitively to students with mild disabilities (Johnson, 1998), as they are typically characterized as having problems with memory and cognition, language, attention, social skills, academic performance, and behavior. Finally, as shown in Table 2, the students with educational, medical, or occupational therapy diagnoses were subdivided into four groups: students with mild disabilities (i.e., ADHD, LD, SI, and/or DCD), students with neurological developmental disorders (e.g., cerebral palsy), students with cognitive disorders (i.e., mental retardation or intellectual disabilities), and students with other or multiple disorders. These categories were used as they are the ones that have been used in the School AMPS manual (Fisher, et al., 2007) to describe the School AMPS standardization sample.

The majority of participants were boys (see Table 2), which likely reflects the higher incidence of boys with reported behavioral

¹ When occupational therapists submit data to the School AMPS database, they are required to indicate up to three most relevant medical, educational, or occupational-therapy-specific diagnoses (including well, typically-developing) for each student. There is no information available to determine by whom or how diagnoses were established for the participants included in the School AMPS database.

problems and/or mild disabilities reported in the literature (Kopp, 2010). We did not consider gender as a variable in any of our studies because prior research has revealed no gender-related differential item or task functioning nor gender differences in mean quality of schoolwork task performance (Fisher, et al., 2007; Sperens, et al., 2010). Gender, therefore, was not considered to be a threat to internal design validity.

Table 2. Age, gender, diagnostic group, and world region for the participants included in Studies I to IV

Study	I	II	III	IV
Number of participants	6194	984	350	36
Age range (yrs)	3 to 21	3 to 13	4 to 11	5 to 10
<i>M (SD)</i>	7.4 (2.8)	6.9 (2.3)	6.9 (1.8)	7.2 (1.6)
Gender				
Boys	4155	688	230	32
Girls	2035	296	120	4
Diagnosis groups				
Typically-developing	1013	148	175	-
At risk	523	68	-	23
Mild	1304	184	175	13
Developmental/neurological	2006	440	-	-
Cognitive/psychiatric	434	36	-	-
Other	914	108	-	-
World regions				
North America	2512	246	180	-
United Kingdom/Republic of Ireland	1042	246	31	-
Other Europe	298	-	7	-
Nordic countries	1255	246	35	36
Australia/New Zealand	572	246	89	-
Central & South America/Caribbean	40	-	1	-
Asia	475	-	7	-

Data extraction procedures

All data were gathered retrospectively. More specifically, data for all potential participants for Studies I, II, and III were retrieved from the School AMPS database at the AMPS Project International, Fort Collins, Colorado, USA, and the data for the potential participants in Study IV were retrieved from the records of an ongoing demonstration project in a medium sized municipality in Sweden. A summary of the *initial* participant extraction and subsequent inclusion or exclusion are shown in Table 3.

Table 3. Number and date of extraction of the potential participants initially extracted from the School AMPS database (Studies I, II, and III) and student records (Study IV)

	Total <i>n</i> initial extraction	Date of extraction	Excluded (<i>n</i>)	Final sample (<i>n</i>) *
Study I	6244	June 2009	50	6194
Study II	4043	January 2008	239	984
Study III	1221	December 2004	871	350
Study IV	69	September 2010	33	36

* See Table 2 for final sample characteristics

Data associated with rater scoring error had been excluded prior to extraction of any data from the School AMPS database. The term *rater scoring error* is used to refer to data from School AMPS raters who overall have demonstrated the ability to score the School AMPS in a reliable and valid manner, but who scored an occasional student too leniently or too strictly (i.e., in an invalid manner). Approximately 4% of the potential sample had been omitted (prior to extraction) for this reason. Further details related to the data excluded due to rater scoring error were not available. Additional details of the data extraction procedures, along with subsequent inclusion and exclusion criteria for each study, are described in more detail in the sections that follow.

Study I

For the purpose of Study I, reliability estimation of the School AMPS measures, we included all students in the School AMPS database who had been evaluated based on performance of at least two School AMPS tasks. In total, there were 6244 students, 3 to 21 years of age, in the School AMPS database at the time of extraction (see Table 4). From this potential sample, we omitted 42 students who had performed only one task. Among the remaining 6202 students, we found 8 students not previously omitted who had unusually high school motor quality of performance measures associated with rater scoring error. These students, therefore, also were removed, resulting in a final sample of 6194 students (see Table 2).

Study II

For the purpose of Study II, cross-regional validation of the School AMPS, all students in the School AMPS database at the time of extraction were selected (see Table 3). In total, 4043 students, 3 to 15 years of age, were then divided into seven groups by world region: North America (United States and Canada, NA), United Kingdom and the Republic of Ireland (UK), Nordic countries, other European countries, Australia and New Zealand (ANZ), Central and South America and Caribbean (CSA), and Asia (Korea, Japan, China). Because there were insufficient data from other European countries ($n=147$), CSA ($n=40$), or Asia ($n=52$), these world regions were excluded. That is, sample sizes of at least 200 persons are recommended when implementing the DIF analyses (Tennant & Pallant, 2007; Wright & Stone, 1979; Zumbo, 1999).

Students from the four remaining world regions ($n = 3804$) were then matched for age and diagnosis. To ensure that we were blind to the student's school motor and school process quality of schoolwork performance measures when creating our matched data set, we used a data file where only the student's age, diagnosis, and world region were included. The final sample comprised of 984 students – 246 students in each world region.

Study III

For the purpose of Study III, comparison of typically-developing students and students with mild disabilities, 1221 participants, ages 3 to 21 years, were initially extracted from the School AMPS

database (see Table 3). The final sample for Study III was then selected from the extracted sample in two phases. First, all potential students who met the following inclusion criteria were selected: (1) were between 4 to 11 years of age; (2) were typically-developing students or students with mild disabilities; and (3) had no other developmental, neurological, cognitive, or psychiatric disorder (e.g., cerebral palsy, mental retardation, autism). There were 401 students who met our inclusion criteria, 209 typically-developing students and 192 students with mild disabilities. In the second phase, these students were then matched for age and diagnostic group (typically-developing vs. mild disabilities). During the matching process, which is described in more detail below, 51 additional students were eliminated, resulting in a final sample of 350 students from diverse world regions (see Table 2).

During the matching process, we again removed the school motor and school process quality of performance measures from the dataset so as to ensure blind matching. Since there were not enough data to match individual participants for age, we initiated the matching process by first dividing the students into 2-year age groups. We then compared each age-matched diagnostic group and randomly omitted students to ensure equal numbers between groups. More specifically, because all of the age groups except the 6- to 7-year-old group had unequal numbers, we randomly removed students from the group with the higher number until all paired age groups had equal numbers. As noted above, 51 students were removed, 31 typically-developing students from the 4- to 5-year-old age group, 17 students with mild disabilities from the 8- to 9-year-old age group, and 3 typically-developing students from the 10- to 11-year-old age group. This resulted in our final sample of 350 students – 175 typically-developing students and 175 students with mild disabilities (see Table 2).

Study IV

In Study IV, evaluation of the sensitivity of the School AMPS quality of performance measures for detecting change after intervention, the potential participants were all students who had been referred for occupational therapy services based on their teachers' concerns with participation in the classroom, and who were subsequently evaluated using the School AMPS. The final inclusion criteria were that the students had (1) been evaluated with the School AMPS at least twice within a 9-month school year, and (2) received occupational therapy

interventions between the two School AMPS evaluations. At the time of the study, 69 students had been evaluated with the School AMPS between May 2006 and April 2009, of whom 36 students met the inclusion criterion (see Tables 2 and 3, and Figure 2).

As shown in Figure 2, almost 50% of the students did not meet the inclusion criteria. The majority of the excluded students who did not meet our inclusion criteria had been evaluated on only one occasion. Among the documented reasons for lack of a second School AMPS evaluation were that the students were determined not to be in need of occupational therapy services, and therefore, were discharged; or the students were in need of further occupational therapy evaluation related to performance of school tasks not represented in the School AMPS (e.g., activities of daily living, play, crafts, gym). A few students who were discharged from occupational therapy were referred to other healthcare professionals for additional evaluation (e.g., psychiatrist). The majority of the other students who were excluded did receive occupational therapy interventions, but the intervention and/or reevaluation process extended into the next school year because of problems scheduling needed meetings with teachers to plan or implement further occupational therapy services, or because of other internal organizational issues within the school setting (e.g., limited time or expressed teacher concerns about heavy workloads).

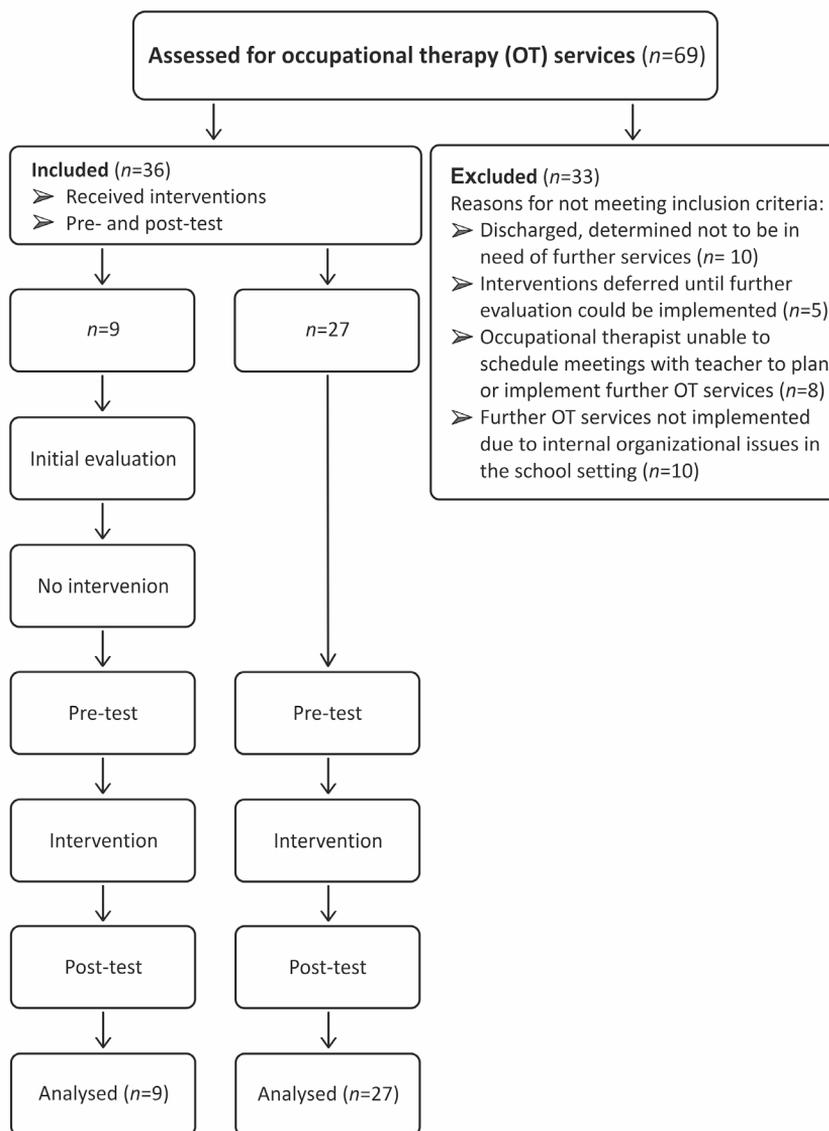


Figure 2. Flow chart over the data extraction, evaluation, and intervention phases for Study IV.

Instrumentation

All participants in Studies I to IV had been evaluated using the School AMPS according to the standardized procedures outlined in the School AMPS manual (Fisher, et al., 2007). The occupational therapists who administered the School AMPS had been trained and calibrated as reliable and valid School AMPS raters. As discussed earlier, the standardized administration procedures specify that the process begins with the occupational therapists interview of the teacher so as to gain an understanding of the student, his or her performance context, and which schoolwork tasks the teacher identifies as presenting a challenge for the student.

During the interview, the occupational therapist also establishes a collaborative working relationship with the teacher. After the interview, the occupational therapist schedules a time to observe the student in his or her natural classroom environment, during normal classroom routines, as the student performs a minimum of two schoolwork task that have been assigned to the students by the teacher. During the observation, the occupational therapist acts as an unobtrusive observer and takes observational notes that he or she later uses to score the quality of the student's schoolwork task performance. When the occupational therapist later scores the quality of the student's observed schoolwork task performance, he or she first matches (1) the characteristics of each of the schoolwork tasks the teacher has assigned to (2) the standardized criteria for the 25 schoolwork tasks included in the School AMPS manual (Fisher, et al., 2007). Assuming that the observed tasks meet the criteria to be classified as being School AMPS tasks, the occupational therapist then scores the 16 school motor and 20 school process skill items for each School AMPS task the student performed. During this process, the occupational therapist matches his or her observations to specific scoring criteria included in the School AMPS manual.

When scoring is complete, the occupational therapist enters the School AMPS task code and the student's school motor and school process skill item raw scores for each task into the School AMPS computer-scoring program (Three Star Press, 2005). Further description of the School AMPS, and the evidence for reliability and validity of the School AMPS scales and quality of schoolwork

performance measures, were discussed in the introduction of this thesis. Therefore, they will not be presented again in this section.

Procedures and data analyses

Many-faceted Rasch analyses

Initially, in all four studies, we implemented MFR analyses, using the FACETS computer program (Linacre, 2009), to generate linearized school motor and school process quality of schoolwork performance measures for each student. We always performed paired MFR analyses – one for the School AMPS motor scale and one for the School AMPS process scale. Just as has been incorporated into the School AMPS computer-scoring program, the majority of the MFR analyses we implemented using FACETS were four-faceted MFR analyses which adjusted the student's final school motor and school process quality of performance measures for the difficulty of the skill items, the challenge of the tasks, and the severity of the rater who scored each student's schoolwork task performances. Thus, these MFR analyses calibrate each student's quality of schoolwork performance measure, the difficulty of each item, the challenge of each task, and the severity of each rater along the same linear scale (school motor or school process).

Although at least four facets were included in all of our analyses (students, items, tasks, and raters), only two of these facets were in focus in our studies – students and items. Therefore, in the various studies implemented in this thesis, we anchored different facets depending on which facet was in focus. Anchoring refers to a procedure that specifies that preset values are to be used in the analysis. In this thesis, those values were those values incorporated into the School AMPS computer-scoring program (Three Star Press, 2005). When a facet is not anchored, new values for each element included in the facet (e.g., individual students, individual items) will be estimated during the MFR analysis (Linacre, 2010). For Studies I, III, and IV, when the student quality of schoolwork performance measures were in focus, all facets except students were anchored. In contrast, in Study II, when the item difficulty calibration values were in focus, we anchored all facets except items. In Study III, when item difficulty calibration values were in focus, we anchored all facets

except items and students (see Table 4); the obtained results for items was identical to what they would have been had we also anchored students.

Table 4. Summary of the number of paired (school motor and school process) MFR analyses performed and which facets were anchored in Studies I to IV

	MFR analyses	Facets anchored
Study I	<u>Five MFR analyses:</u> <ul style="list-style-type: none"> • Student measures based on Task 1 • Student measures based on Task 2 • Student measures based on even items both tasks • Student measures based on odd items both tasks • Student measures based on all items, both tasks 	<ul style="list-style-type: none"> • All facets except students
Study II	<u>Five MFR analysis:</u> * <ul style="list-style-type: none"> • Item DIF calibration values with region added as a fifth facet • Student measures based on NA item calibration values • Student measures based on UK item calibration values • Student measures based on Nordic item calibration values • Student measures based on ANZ item calibration values 	<ul style="list-style-type: none"> • All facets except items; all four regions were anchored at zero • All facets except items
Study III	<u>Three MFR analyses:</u> <ul style="list-style-type: none"> • Student measures, all participants • Item difficulty calibrations based on typically-developing students • Item difficulty calibrations based on students with mild disabilities 	<ul style="list-style-type: none"> • All facets except students • All facets except items and students
Study IV	<u>One MFR analysis:</u> <ul style="list-style-type: none"> • Students measures based on two tasks for each School AMPS evaluation 	<ul style="list-style-type: none"> • All facets except students

* NA = North America, UK = United Kingdom and Republic of Ireland, Nordic = Nordic countries, ANZ = Australia and New Zealand

Descriptive and inferential statistical analyses

We also used the Statistical Package for the Social Sciences (SPSS) to perform descriptive and inferential statistical analyses in all of our studies. Additional descriptive and inferential statistics used in our studies were generated by FACETS. A summary of the descriptive and inferential statistical analyses we implemented is shown in Table 5. Additional details of procedures and implemented data analyses for each study are described in more detail in the following sections.

Table 5. Summary of the descriptive and inferential analyses performed in Studies I to IV

	Statistical analysis	Program
Study I	<ul style="list-style-type: none"> • <i>M</i> and <i>SD</i> • Pearson product moment correlation (<i>r</i>) • Spearman–Brown ($r=2(r)/1+r$) • Rasch equivalent of Cronbach’s alpha (<i>R</i>) • Overall (mean) <i>SEm</i> 	<ul style="list-style-type: none"> • SPSS • SPSS • Calculated by hand • FACETS • FACETS
Study II	<ul style="list-style-type: none"> • <i>M</i> and <i>SD</i> • Standardized difference (<i>Z</i>) using normalized <i>SEm</i> [$Z=D_1-D_2/\sqrt{SEm_1^2+SEm_2^2}$] • Plots of paired world region quality of performance measures (DTF) 	<ul style="list-style-type: none"> • SPSS • SPSS • SPSS
Study III	<ul style="list-style-type: none"> • <i>M</i> and <i>SD</i> • Independent samples <i>t</i> tests • Relative difference in skill item difficulty calibration values (D_1-D_2) • Actual differences in skill item difficulty calibration values (relative difference + group <i>M</i> difference) • Standardized difference (<i>Z</i>) [$Z=D_1-D_2/\sqrt{SEm_1^2+SEm_2^2}$] 	<ul style="list-style-type: none"> • SPSS • SPSS • SPSS • SPSS • SPSS
Study IV	<ul style="list-style-type: none"> • <i>M</i> and <i>SD</i> • Percent • Repeated measures ANOVAs (Total sample, Group 1) • Post hoc <i>t</i> tests (Group 1) • Paired samples <i>t</i> tests (Group 2) 	<ul style="list-style-type: none"> • SPSS • SPSS • SPSS • SPSS • SPSS

Study I

Evidence of reliability

In the first and second sets of paired MFR analyses implemented for Study 1 (see Table 4), we split our data “vertically” and generated student quality of schoolwork performance measures based on all skill items for Task1 and based on all items for Task2. In the third and fourth sets of paired MFR analyses, we split our data “horizontally” and generated student quality of performance measures based on half the School AMPS items (odd and even) (see Table 4 and Figure 3). We then used Pearson product moment correlation (r) analyses to evaluate the consistency of the student quality of schoolwork performance measures between Task1 and Task2 (i.e., “vertical” split-half coefficient of equivalence), and between odd and even skill items (i.e., “horizontal” split-half coefficient of equivalence), respectively. Since the reliability estimates were based on only half of a full School AMPS evaluation, we adjusted the estimated r from each split-half analysis using the Spearman–Brown formula (Ary, et al., 2002; Crocker & Algina, 1986). We set our criterion for a minimum acceptable reliability coefficient at ≥ 0.70 (Cohen & Swerdlik, 2010).

In the final set of paired MFR analyses implemented for Study I (see Table 4), we used FACETS to generate reliability statistics in the form of (1) the Rasch equivalent of Cronbach’s alpha and (2) the overall error variance (the overall SEm , based on the mean of the SEm s estimated for each student included in the analyses). These reliability estimates were used to cross-validate our “vertical” and “horizontal” split-half reliability estimates described above. We also compared the MFR generated overall SEm from the four vertical and horizontal split-half analyses to the overall SEm generated from the final set of MFR analyses based on a full School AMPS evaluation (all items and two tasks).

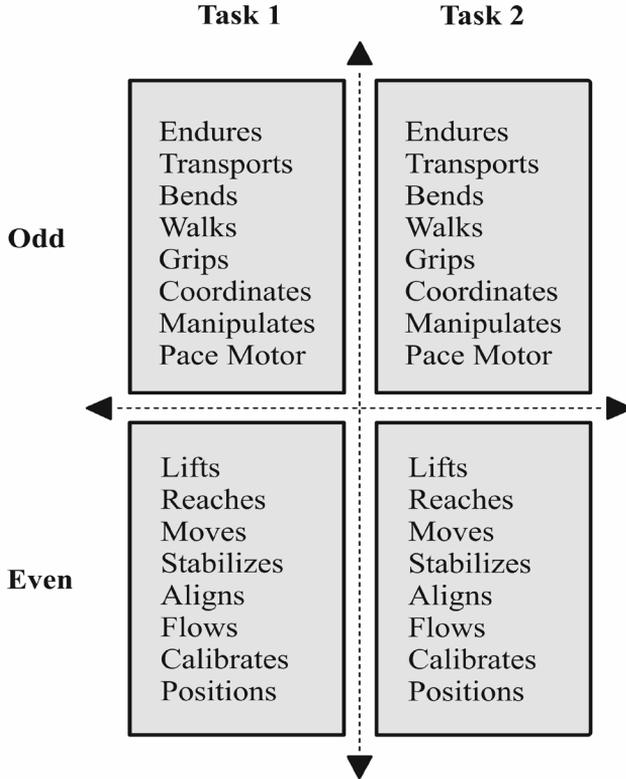


Figure 3. Illustration of how the school motor skill items from one school AMPS evaluation were split “vertical” and “horizontal” into two halves.

In the final set of paired MFR analyses implemented for Study I (see Table 4), we used FACETS to generate reliability statistics in the form of (1) the Rasch equivalent of Cronbach’s alpha and (2) the overall error variance (the overall *SEm*, based on the mean of the *SEms* estimated for each student included in the analyses). These reliability estimates were used to cross-validate our “vertical” and “horizontal” split-half reliability estimates described above. We also compared the MFR generated overall *SEm* from the four vertical and horizontal split-half analyses to the overall *SEm* generated from the final set of MFR analyses based on a full School AMPS evaluation (all items and two tasks).

Study II

Evidence of validity based on internal structure

In the first DIF comparison, each world region was compared to each of the others as illustrated in Figure 4. Since there are 16 school motor skill items and 20 school process skill items, there were six comparisons for each item, resulting 96 school motor and 120 school process comparisons, respectively.

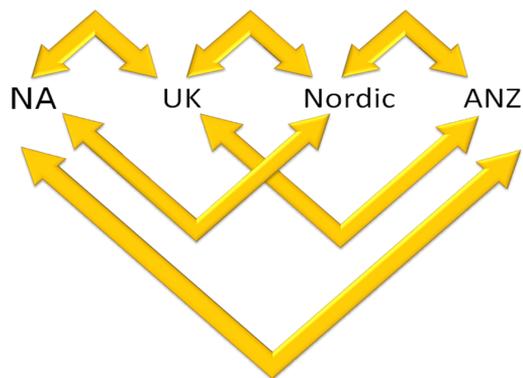


Figure 4. DIF comparisons where each world region skill item difficulty calibration value was compared to the skill item difficulty calibration values for each of the other three world regions. (**Note.** NA = North America, UK = United Kingdom and Republic of Ireland, Nordic = Nordic countries, ANZ = Australia and New Zealand)

In the second DIF comparison (see Figure 5), the skill item difficulty calibrations values for each world region (e.g., NA) were compared to the skill item difficulty calibration values for the total sample (i.e., all four regions combined). In total, four comparisons were made, one for each world region. We set our criteria for significant DIF at ± 0.55 logit (Tristán, 2006).

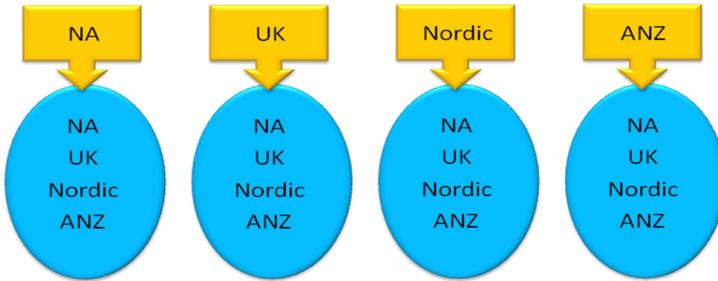


Figure 5. DIF comparisons where each world region skill item difficulty calibration values were compared to the skill item difficulty calibration values for the total sample (all four world regions combined). (**Note.** NA = North America, UK = United Kingdom and Republic of Ireland, Nordic = Nordic countries, ANZ = Australia and New Zealand)

Evidence of validity related to consequences of testing (DTF)

As summarized in Table 4 and shown graphically in Figure 6, we implemented four different sets of paired MFR analyses (i.e., four for school motor and four for school process) to generate new sets of student quality of schoolwork performance measures based on the item difficulty calibration values for each world region that had been generated by the first set of paired MFR analyses for DIF. For example, in the first set of these MFR analyses, all students' ordinal school motor and school process skill item raw scores were converted to linear quality of schoolwork performance measures based on the NA skill item difficulty calibration values.

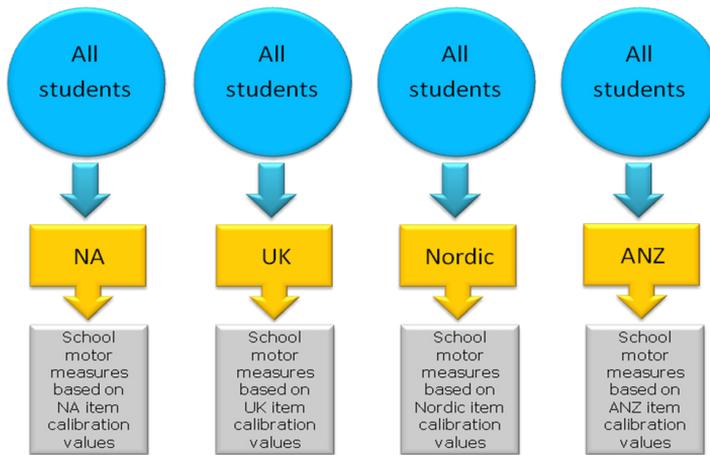


Figure 6. Four sets of student quality of schoolwork performance measures generated based on four different sets of world region skill item difficulty calibration values. (**Note.** NA = North America, UK = United Kingdom and Republic of Ireland, Nordic = Nordic countries, ANZ = Australia and New Zealand)

In the next step (see Table 5), analyses for DTF were performed by plotting students' quality of schoolwork performance measures based on one world region against students' quality of schoolwork performance measures from another world region. For example, we plotted the students' quality of performance measures based on the NA item difficulty calibration values against the students' quality of performance measures based on the ANZ item difficulty calibration values. This resulted in 12 comparisons, 6 for school motor and 6 for school process. We then constructed 95% confidence bands based on the sum of the *SEms* for each of the student's paired quality of performance measures. Thus, we set our criteria for no DTF if at least 95% of the paired measures fell inside the confidence bands (Wright & Masters, 1982).

Study III

In the first of the three sets of paired MFR analyses performed for Study III (see Table 4), we generated quality of performance measures for all students included in our sample. In the second and third sets of paired MFR analyses, we performed separate analyses to generate skill item difficulty calibration values for the typically-developing students and for the students with mild disabilities.

Evidence of validity based relations to other variables (sensitivity)

Initially, independent samples *t* tests were performed to examine for group differences in mean quality of schoolwork task performance between typically-developing students and students with mild disabilities. We set the level of significance at $p < 0.05$.

Evidence of validity based on internal structure (DIF)

Then, we proceeded to calculate the *relative* differences in skill item difficulty calibration values by calculating the difference between those based on the typically-developing students and those based on the students with mild disabilities. Since the majority of the skill item difficulty calibration values were associated with low *SEms* (< 0.10 logit), we set the criterion for a clinically meaningful difference at ≥ 0.43 logit. This criterion is equivalent to a 95% confidence interval when the *SEm* for the skill item difficulty calibration values is 0.15 logit (Bernspång & Fisher, 1995; Cooke, Fisher, Mayberry, & Oakley, 2000; Duran & Fisher, 1996; Fisher, Goldman & Fisher, 1997; Magalhães, Fisher, Bernspång, & Linacre, 1996; Fisher, Kilgore, & Harley, 1992; Oakley, Duran, Fisher, & Merritt, 2003; Silverstein, 1992).² To further evaluate for statistical differences in relative item difficulty calibration values, we also computed the standardized difference (*Z*) (Wright & Masters, 1982; Wright & Stone, 1979).

² In Study II, which was implemented chronologically later than was Study III, we used a less strict criterion of 0.55 logit, which we based on Tristán (Tristán, 2006). More specifically, Tristán has determined that the smallest possible *SEm* is 0.20 logit, which is larger than the 0.15 logit criterion we had used earlier in Study III.

Evidence of validity based on consequences of testing

Our final step was to calculate for *actual* differences in skill item difficulty calibrations between groups. This was done by adding the group mean difference in school motor and school process quality of schoolwork task performance to the relative difference in skill item difficulty calibration values to adjust these relative differences to account for any differences in quality of schoolwork task performance between groups (Kottorp, Bernspång, & Fisher, 2003). As with the evaluation of the relative differences, we set the criteria for meaningful differences in actual skill item difficulty calibration values at ≥ 0.43 logit.

Study IV*Evidence of validity based on relations to other variables (sensitivity)*

In Study IV, we implemented our data analyses in three phases. In the first phase, we implemented a 2 X 2 (special education X time) repeated measures ANOVA based on all 36 students to assess whether or not the receipt of special educational services had a significant effect on the students' mean quality of schoolwork task performance over time.

We then proceeded with the second phase and carried out several additional analyses. First, we performed two one-way repeated measures ANOVAs to evaluate for significant improvements in mean school motor and school process quality of schoolwork task performance for the 9 students evaluated three times (Group 1) (see Figure 2). Second, we implemented two paired samples *t* tests, one for school motor and one for school process quality of performance, to evaluate for significant differences in mean pre- and post-test school quality of performance measures among the 26 students (Group 2) who had no control phase. For the 9 students in Group 1, we also implemented post hoc *t* tests to identify where significant improvements occurred (i.e., between pre-test1 and post-test, between pre-test2 and post-test, between pre-test1 and pre-test2). Finally, we calculated effect sizes using partial η^2 (Cohen, 1988) for Group 1 and the formula $ES = M_1 - M_2 / SD_1$ (Becker, 2000) for Group 2. Both effect sizes were interpreted based on Cohen's (1988) criteria.

In the third phase, we determined how many of the 36 students demonstrated a clinically meaningful difference in school motor and/or school process quality of performance measures pre- and post-intervention by calculating the difference between each student's paired school motor and school process quality of performance measures. Our criterion for a clinically meaningful change was set to ≥ 0.30 logit, based on the criteria included in the School AMPS manual (Fisher, et al., 2007). We also evaluated for statistically significant differences in quality of schoolwork task performance post-intervention for each of the 36 students based on criteria recommended by Harvill (1991) (i.e., differences between paired measures of at least 2 *SEm*). Additionally, when going through the records for participants in Study IV, we were able to determine number of collaborative, contacts and types of interventions implemented during the intervention phase.

Sample size and effect size considerations

In all of our analyses, we attempted to use the largest possible sample, and a priori power analyses were not performed. More specifically, in Studies I and IV, all available data were used, and in Studies II and III, all available data that could be matched were used. For our DIF analyses in Study II, we considered existing recommendations where sample sizes of at least 200 participants in each group is preferred (Tennant & Pallant, 2007; Zumbo, 1999). In Study III, we had only 175 students in each group. According to Linacre (1994), sample sizes of 100 to 144 per group are needed to ensure 95% confidence for stable skill item difficulty estimates based on a criterion of ± 0.50 logit.

In Study II, effect sizes for evaluating significant DIF were based on Tristán (2006). In Study III, post-hoc effect sizes (*d*) were calculated based on the standardized mean difference (Cohen, 1988). Similarly, in Study IV, we calculated post-hoc effect sizes based on partial *Eta*² and Cohen's *d* (1988). The magnitudes of *d* and partial *Eta*² were interpreted based on the guidelines proposed by Cohen (1988).

Ethical considerations

Approval for the studies included in this thesis was obtained from the Ethical Committee of the Medical Faculty, Umeå University,

Sweden (Study I, II, and III, Dnr 03-509 and 06-051M; Study IV, Dnr 09-081M).

Results

Evidence of reliability and validity of the School AMPS measures and scales

Evidence of reliability

As shown in Table 6, the two vertical forms and the two horizontal forms had virtually equivalent means, standard deviations (*SD*), and standard errors of measurement (*SEm*). This indicates that we did have strictly parallel forms (Feldt & Brennan, 1989).

Table 6. Means, standard deviations (*SD*), and standard errors of measurement (*SEm*) for each split-half of the School AMPS (Study I)

	School motor			School process		
	Mean	<i>SD</i>	<i>SEm</i>	Mean	<i>SD</i>	<i>SEm</i>
“Vertical” split						
Task 1	1.60	0.78	0.41	0.44	0.66	0.31
Task 2	1.62	0.77	0.41	0.46	0.64	0.31
“Horizontal” split						
Odd	1.46	0.85	0.42	0.46	0.65	0.31
Even	1.52	0.80	0.39	0.42	0.61	0.32

As shown in Table 7, the Spearman–Brown corrected split-half reliability estimates were high ($r \geq 0.70$), and all but one were $r \geq 0.80$. The Rasch equivalents of Cronbach’s alpha also were high, $R = 0.85$ for both the school motor and school process quality of schoolwork performance measures estimated based on a full School AMPS evaluation (all items, two schoolwork tasks). Finally, the overall mean *SEm*s were 0.28 logit for school motor and 0.22 logit for school process.

Table 7. Reliability estimates for the school motor and school process quality of performance measures (Study I)

School AMPS (full evaluation)	School motor	School process
Corrected r (vertical)	0.86	0.73
Corrected r (horizontal)	0.82	0.82
Rasch equivalent of Cronbach's alpha	0.85	0.85
Overall SEm	0.28	0.22

Evidence of validity based on internal structure and consequences of testing

Stability of the skill item hierarchies (DIF)

Overall we found the hierarchy of the 16 school motor and 20 school process skill items to be stable between groups of students from different world regions (Study II), and when comparing typically-developing students and students with mild disabilities (Study III).

In Study II, when we compared the school motor and school process skill item hierarchies for each individual world region to the skill item hierarchies for each of the other regions, we found that three school motor skill items (Walks, Moves, and Endures), and one school process skill item (Navigates) displayed DIF (see Table 8). For example, the skill item Walks was significantly easier to perform for students from North America (NA) than for students from Australia and New Zealand (ANZ), as indicated by the positive difference ≥ 0.55 logit.

Table 8. Items displaying DIF among world regions

Skill items	Region comparisons	Measure	SEm	Measure	SEm	DIF*
School motor skill item difficulty calibration values						
Walks	NA vs. ANZ	1.15	0.08	0.53	0.06	0.62
Moves	NA vs. UK	1.27	0.09	0.67	0.06	0.61
Moves	UK vs. ANZ	2.37	0.12	1.81	0.10	0.56
Endures	UK vs. Nordic	0.67	0.06	1.24	0.08	-0.58
School process skill item difficulty calibration values						
Navigates	UK vs. Nordic	0.36	0.05	0.92	0.05	-0.56

Note. NA = North America, UK = United Kingdom and Republic of Ireland, Nordic = Nordic countries, ANZ = Australia and New Zealand

In the second DIF analyses, however, when the skill item hierarchies for each individual world region were compared to the skill item hierarchies for the total sample that included all four world regions (see Figure 5), none of the skill items displayed DIF. Thus, the results from this analysis indicate that the skill item hierarchies for both school motor and school process remained stable in relation to the overall item calibration values.

Because the presence of DIF can be an indication of unconfirmed group differences in item performance among these world regions (Pennfield & Camilli, 2007) and/or suggest a secondary dimension in the School AMPS scales (Osterlind & Everson, 2009), we analyzed our data further to see if we could identify any possible sources of the DIF among world regions. We found no systematic differences in item calibration values among regions that could be associated with gender, specific diagnosis, tasks performed, or rater.

In Study III, where we compared the item calibration hierarchies for the typically-developing students to the item calibration hierarchies for the students with mild disabilities, we found that four school motor skill items displayed DIF (Walks, Transports, Lifts, and Endures), but that the school process skill item hierarchy was stable between the two groups (see Tables 9 and 10). Since these four

school motor skill items were among the easiest actions to perform skillfully, especially for students without obvious physical, cognitive, or neurological disabilities, we suspected estimation error due to off-targeting as a source of disruption. We, therefore, computed standardized difference (Z), and found that only one school motor item (Endures) demonstrated statistically significant DIF ($p \leq .05$).

Table 9. School motor skill item difficulty calibrations (logits) for typically-developing students and students with mild disabilities (Study III)

Item	Typically-developing	Mild disabilities	Relative difference	Actual difference
Transports	2.03	1.45	0.58*	1.07*†
Lifts	1.82	1.27	0.55*	1.04*†
Walks	1.70	0.97	0.73*	1.22*†
Moves	1.45	1.04	0.41	0.90*†
Bends	1.23	1.35	-0.12	0.37
Reaches	1.06	1.35	-0.29	0.20
Endures	1.00	1.79	-0.79*	-0.30
Stabilizes	-0.35	0.06	-0.41	0.08
Grips	-0.87	-0.61	-0.26	0.23
Coordinates	-0.91	-0.62	-0.29	0.20
Aligns	-0.96	-0.91	-0.05	0.44*
Flows	-1.04	-0.85	-0.19	0.30
Manipulates	-1.36	-1.11	-0.25	0.24
Calibrates	-1.43	-1.32	-0.11	0.38
Paces	-1.43	-1.83	0.40	0.89*†
Positions	-1.93	-2.00	0.07	0.56*†

* Represents a meaningful difference ≥ 0.43 logit

† Items that differed significantly in actual challenge ≥ 0.55 logit

Table 10. School process skill item difficulty calibrations (logits) for typically-developing students and students with mild disabilities (Study III)

Item	Typically-developing	Mild disabilities	Relative difference	Actual difference
Chooses	1.13	1.23	-0.10	0.48*
Searches/Locates	1.05	1.13	-0.08	0.50*
Adjusts	0.74	0.54	0.20	0.78*†
Gathers	0.73	0.84	-0.11	0.47*
Sequences	0.70	0.74	-0.04	0.54*
Inquires	0.58	0.60	-0.02	0.56*†
Uses	0.52	0.77	-0.25	0.33
Navigates	0.47	0.50	-0.03	0.55*†
Terminates	0.09	-0.05	0.14	0.72*†
Heeds	0.07	-0.27	0.34	0.92*†
Restores	0.03	0.03	0.00	0.58*†
Initiates	-0.34	-0.40	0.06	0.64*†
Handles	-0.37	-0.07	-0.30	0.28
Paces	-0.39	-0.49	0.10	0.68*†
Organizes	-0.39	-0.11	-0.28	0.30
Notices/Responds	-0.54	-0.73	0.19	0.77*†
Continues	-0.71	-0.55	-0.16	0.42*
Attends	-0.96	-0.83	-0.13	0.45*
Benefits	-1.00	-1.15	0.15	0.73*†
Accommodates	-1.41	-1.72	0.31	0.89*†

* Represents a meaningful difference ≥ 0.43 logit

† Items that differed significantly in actual challenge ≥ 0.55 logit

Differential test functioning (DTF)

Further evaluation for *evidence related to consequences of testing* revealed the skill items displaying DIF were not impacting the School AMPS measurement system for students in the four world regions included in Study II. That is, when the students’ quality of schoolwork performance measures generated based on different world region skill item difficulty calibration values were compared to each other (see Figure 7), no evidence of DTF was found.

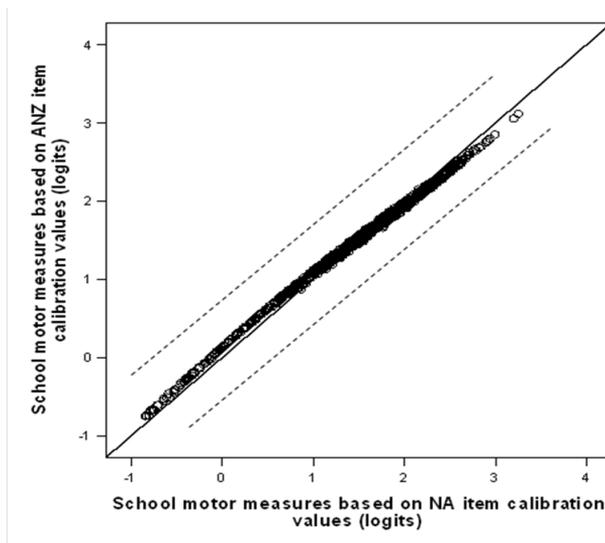


Figure 7. Scatterplot of the students' school motor quality of performance measures based on ANZ skill item calibration values versus the students' school motor quality of performance measures based on the NA skill item calibration values. All paired measures fell within the 95% confidence interval.

Identification of actual differences in skill challenge

We also considered our evaluation of clinically meaningful differences in actual skill item difficulties to be related to *evidence of consequences of testing* (Study III). Our rationale was that knowing which skills students with mild disabilities actually have a greater challenge performing could be of future benefit when planning intervention programs. We found that the students with mild disabilities demonstrated greater actual challenges performing 7/16 (43.8%) of the school motor items, and 17/20 (85%) of the school process skill items than did the typically-developing students in our sample (see Tables 9 and 10). When we used the stricter criteria of ≥ 0.55 logit we used in Study II (Tristán, 2006) as our criteria for a significant (vs. meaningful) difference, we found that the students with mild disabilities actually demonstrated significantly lower quality of performance on 6/16 (37.6%) school motor skills and 11/20 (55%) school process skills.

Evidence based on relations to other variables

Sensitivity when discriminating between typically-developing students and students with mild disabilities

The results from the independent *t* tests comparing the mean quality of schoolwork performance measures for typically-developing students to the mean quality of schoolwork performance measures for students with mild disabilities (Study III) revealed that the students with mild disabilities had significantly lower school motor and school process ability measures than did their typically-developing peers (see Table 11). We concluded, therefore, that the School AMPS measures are sensitive enough to discriminate between groups expected to differ (i.e., typically-developing students and students with mild disabilities).

Table 11. Mean school motor and school process quality of schoolwork performance measures for the typically-developing students and students with mild disabilities (Study III)

Measurement	<i>M</i>	<i>SD</i>	Mean difference	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
School motor (logits)							
Typically developing	2.25	0.50					
Mild disabilities	1.76	0.50	0.49*	9.26	348	<.001	0.89
School process (logits)							
Typically developing	0.98	0.44					
Mild disabilities	0.40	0.46	0.58*	12.04	348	<.001	1.09

* Also represents a clinically meaningful difference ≥ 0.30 logit

Sensitivity when measuring change before and after intervention

In Study IV, 20 students had received special education services, but no significant time by special education interaction or main effects were found for either school motor ($F[1,34] \leq 1.04$; $p \geq .31$) or school process ($F[1,34] \leq 1.75$; $p \geq .20$) quality of performance measures. We proceeded, therefore, to evaluate for significant changes over time,

pre- versus post-intervention. The one-way ANOVA for the 9 students in Group 1 revealed significant changes over time for both school motor and school process quality of performance measures (see Table 12). The post hoc *t* tests revealed significant improvements ($p \leq .05$) in both school motor and school process quality of schoolwork task performance between both pre-test1 and post-test and between pre-test2 and post-test, and there were no significant changes between pre-test1 and pre-test2 (the control phase). The intervention effect size was large (see Table 12). Similarly, the paired sample *t* tests performed for Group 2 revealed significant improvements in both mean school motor and mean school process quality of schoolwork task performance between the pre-test and the post-test evaluations; the effect sizes were again large (see Table 13). A comparison of the improvements in school motor and school process ability measures for Group 1 and Group 2 are presented graphically in Figures 8 and 9. Based on our results, we concluded that the School AMPS measures are sensitive indices of change post interventions.

Table 12. Results of repeated measures ANOVA test within subjects effects and post hoc *t* test for Group 1 (Study IV)

Group 1 (n=9)	<i>M</i>	<i>SD</i>	Mean difference	df	<i>F</i>	<i>p</i>	Partial η^2
School motor (logits)							
Pre-test1	1.58	0.31					
Pre-test2	1.68	0.23	0.10			0.22	
Post-test	2.07	0.31	0.39*	2, 16	10.82	<.001	0.58
School process (logits)							
Pre-test1	0.21	0.44					
Pre-test2	0.14	0.36	-0.07			0.53	
Post-test	0.62	0.36	0.48*	2, 16	17.46	<.001	0.69

* Also represents a clinically meaningful difference ≥ 0.30 logit

Table 13. Results of paired sample *t* test for students in Group 2 (Study IV)

Group 2 (<i>n</i> =27)	<i>M</i>	<i>SD</i>	Mean difference	<i>t</i>	<i>df</i>	<i>p</i>	<i>ES</i>
School motor (logits)							
Pre-test	1.61	0.30					
Post-test	1.93	0.33	0.32*	4.64	26	<.001	1.02
School process (logits)							
Pre-test	0.25	0.41					
Post-test	0.62	0.39	0.37*	4.33	26	<.001	0.93

* Also represents a clinically meaningful difference ≥ 0.30 logit

When we calculated the differences of the school motor and school process quality of schoolwork performance following intervention for each student (*n*=36) to evaluate for clinically meaningful (i.e., observable) changes, our result revealed that 80.5% of the students improved by at least 0.30 logit in school motor and/or school process quality of schoolwork task performance. We found no pattern suggesting an association between age, length of time between pre- and post-test, receipt of special education services, number of collaborative contacts, or the presence or absence of clinically meaningful change (Table 14). When we used Harvill’s (1991) guideline, and evaluated how many of the students had School AMPS measures that differed by at least the sum of the *SEMs* for each of their paired measures (the *SEM* values were based on the results of Study I), we found that two thirds (66.7%) likely demonstrated significant differences in school motor and/or school process quality of performance.

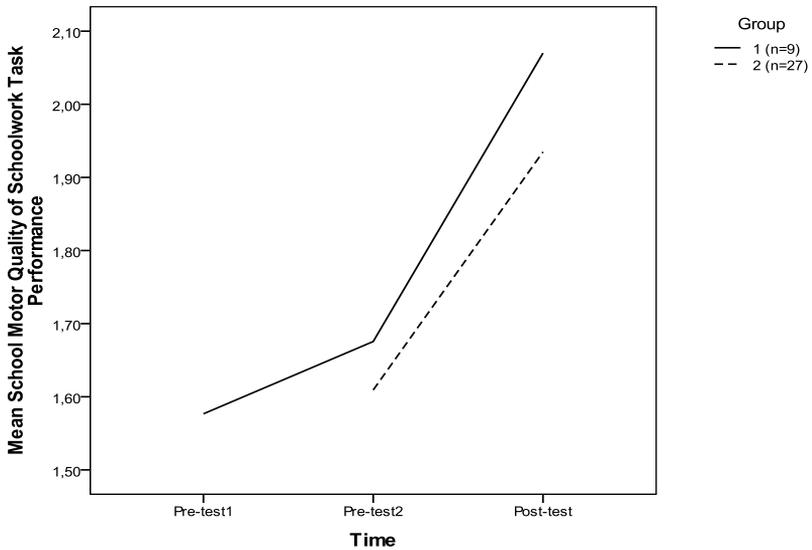


Figure 8. Mean school motor quality of performance measures for Group 1 (pre-, pre-, and post-test) and Group 2 (pre- and post-test only).

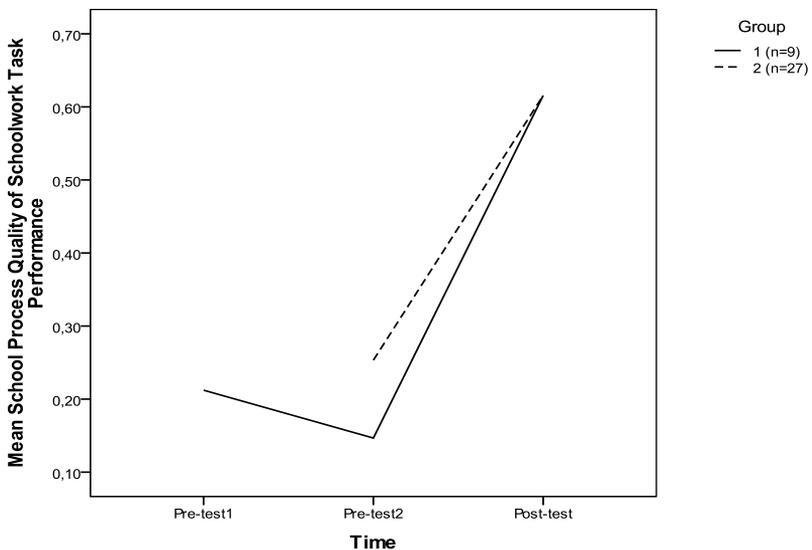


Figure 9. Mean school process quality of performance measures for Group 1 (pre-, pre-, and post-test) and Group 2 (pre- and post-test only).

Description of occupational therapy services

The number of collaborative contacts between the occupational therapist and the teacher varied from 3 to 10 times (overall $M=5.1$) during the course of a school year, including contacts related to planning and implementing the School AMPS observations. The occupational therapists spent most of their time providing consultative services with the teacher related to planning and implementing adaptive occupation (modifying the physical and social environments, teaching new methods of task performance, and provision of adaptive equipment) to compensate for decreased occupational performance skills (see Table 15) (Fisher, 2009). In most cases, the actual interventions were implemented by the teacher. Occasionally, when the occupational therapists entered the classroom, she demonstrated techniques and/or shared knowledge related directly to the student's performance of schoolwork tasks. There were also occasions where the occupational therapist discussed schoolwork-related strategies with the students' parents that the parents were encouraged to implement at home.

Table 15. Types of compensatory interventions implemented during the intervention phase

Provide adaptive equipment	Teach alternative or compensatory strategies	Modify task, or physical or social environments
<ul style="list-style-type: none"> • Use of visual time management strategies (e.g., clock, timer) • Adaptive scissors • Modified seating • Use of computer to write • Use of "fidgets" 	<ul style="list-style-type: none"> • Adaptive strategies to support writing • Individualized working schedules (e.g., use of pictures to clarify what the student is to do or what materials the student is to use, incorporating short pauses into students schoolwork schedule) • Use of stickers as reminders where to stop • Task accommodations (e.g., modified task instructions) 	<ul style="list-style-type: none"> • Added structure in classroom • Enhanced routines in classroom • Revised student placement in classroom • Revised arrangement of furniture in classroom • Modify equipment/tools • Task modifications (i.e., simplified task requirements)

Discussion

While the primary aim for this thesis was to evaluate aspects of reliability and validity of the School AMPS scales and quality of performance measures, my long-term objective was to gather evidence related to the clinical usefulness of the School AMPS measures when used to evaluate students' quality of occupational performance in school settings. When the School AMPS is used to measure students' quality of schoolwork task performances in natural classroom environments, clinical usefulness pertains to such issues as whether or not the occupational therapist can use the results of a School AMPS observation to identify students who are in need of occupational therapy or other related services, plan occupation-based interventions, and evaluate change and document the effectiveness of such interventions. Another issue related to clinical usefulness that must always be considered is whether or not an evaluation tool such as the School AMPS, which was developed in North America, can be used in new world regions (e.g., Nordic countries) without risk of unfair test practices or bias.

While each of these issues ultimately pertain to the validity of the School AMPS measures and scales, the occupational therapist cannot make valid interpretations of a student's quality of schoolwork performance measures if the School AMPS measures are not reliable (Cook & Beckman, 2006). Conversely, even if the School AMPS measures are reliable, this does not imply evidence of their validity. Rather, both reliability and validity need to be evaluated and considered when making decisions about whether or not to use an evaluation tool in clinical practice or research (Downing, 2003).

Therefore, in the following sections, I will discuss first the reliability evidence for the School AMPS measures that was in focus in Study I. Then I will discuss the various aspects of validity evidence for the School AMPS scales and quality of performance measures that were in focus in Studies II to IV. Because of the close tie between (1) reliability and validity and (2) the clinical usefulness of the School AMPS measures in practice (including research), I have chosen to organize the discussion of our results such that the limitations, clinical implications, and recommendations for further research are addressed together in relation to each type of evidence that was in

focus in this thesis. I will then summarize the key clinical implications and main conclusions again at the end of the discussion.

Evidence of reliability

When we initiated the research that is reported in this thesis, the only estimates of reliability that previously had been reported were related to rater reliability. More specifically, more than 90% of the trained and calibrated School AMPS raters have been shown to demonstrate acceptable goodness-of-fit to the MFR model of the School AMPS, supporting overall high inter- and intra-rater reliability (Fisher, et al., 2007). We, therefore, felt it important to continue the process of examining evidence for reliability by determining the reliability of the school motor and school process quality of performance measures. Given that we had available data for a large number of students ($n=6194$) who had been tested using the School AMPS on one occasion only, we chose to use a combination of available methods to evaluate the extent to which the School AMPS measures remain consistent and free from random error. We judged this to be a priority as the more measurement error that exists in the School AMPS measures, the less useful they will be clinically, for example when interpreting a student's School AMPS results or evaluating if his or her quality of schoolwork task performance has improved following intervention (Henson, 2001).

We obtained high reliability coefficient estimates ($r \geq 0.73$) which we cross-validated using three different methods. Thus, we obtained good evidence that the School AMPS measures are consistent. This result was further supported by our low obtained *SE*ms. Given that our results were based on a single School AMPS evaluation, our results suggest that occupational therapists using the School AMPS can feel confident that a student's school motor and school process quality of performance measures would be very similar had he or she been evaluated again under identical conditions (e.g., same time, same place).

The results of the vertical split-half analysis (i.e., Task1 vs. Task2) also suggest that even if the student performs different School AMPS tasks, his or her School AMPS results will be similar, again, given that the student is evaluated under identical conditions. Finally, the

low *SEms* have the potential to be used clinically in the process of clarifying the range within which a student's true measure likely falls (i.e., a student's true measure would likely fall within a range delineated by his or her observed (actual) measure ± 2 *SEm*) (Cohen & Swerdlik, 2010; Crocker & Algina, 1986).

Our results also suggest that the School AMPS measures likely would be sensitive measures of change. That is, Harvill (1991) stated that a difference between two measures that is

equal to or greater than two [SEms] is a meaningful one. A better statement concerning the two [SEms] difference between scores would be the following: 'It is quite likely that the difference between the two scores is a real one and is not likely to have happened by chance alone.' In other words, the difference between the two scores is statistically significant. (p. 186)

Thus, according to Harvill, if two student School AMPS measures differ by at least 2 *SEms*, there is a reasonable chance that the two measures differ significantly. More specifically, if one uses the criterion of 2 *SEm*, then there is an 85% chance the two measures differs significantly (Dawson & Trapp, 2004).

We did find that the estimated mean *SEms* for a full School AMPS evaluation is likely to apply for the majority of our sample. Nevertheless, when applying Harvill's (1991) strategy in practice, the occupational therapist should consider using the sum of the *SEms* for each School AMPS measure (i.e., $SEm_1 + SEm_2$), as the magnitude of the *SEm* does vary depending upon where along the School AMPS scales the student's measures are located (Harvill, 1991; Wright, 1995). I also recognize that a more conservative approach to evaluating for significant differences between two School AMPS measures would be to calculate the standardized difference (Z) (Cohen & Swerdlik, 2010; Wright & Masters, 1982; Wright & Stone, 1979) which we did use in Study III. Harvill (1991) also acknowledges that this method is "more precise" (p. 186).

One concern with the split-half method we used (i.e., splitting the School AMPS items in half vertically or horizontally) is that this procedure results in reducing the length of the test, and therefore

underestimating the reliability coefficient. We addressed this issue by using the Spearman–Brown formula to correct our split-half reliability estimates (Cohen & Swerdlik, 2010). However, this method has been criticized because the size of the reliability coefficient may differ depending on how the items in the instrument are divided, especially when the split halves are not strictly parallel (Becker, 2000; Cohen & Swerdlik, 2010; Streiner, 2003). Given that we were able to confirm that we did have strictly parallel forms (i.e., equivalent means, standard deviations, and standard errors), there was likely no such risk for systematic differences between our vertical and horizontal halves (Allen & Yen, 2002; Crocker & Algina, 1986; Haertel, 2006). Moreover, we were able to cross-validate our split-half method results by comparing them to the Rasch equivalent of Cronbach’s alpha for a “full length” School AMPS evaluation, based on all items and two tasks.

As would be expected, we found that the overall *SEms* generated by our MFR analyses based on the full School AMPS evaluation were lower than were the overall *SEms* based on when tasks and items were split. Clinically, this finding supports the importance of having the student perform at least two schoolwork tasks when he or she is evaluated using the School AMPS. That is, the standardized School AMPS administration procedures require that students perform at least two School AMPS tasks. The results of our study provide evidence to support this practice – having students perform two School AMPS tasks will result in lower *SEms* which, in turn, will enhance the occupational therapist’s possibility to be able to detect significant changes in a student’s quality of schoolwork task performance after providing interventions.

While our use of several different methods for estimating reliability provided good evidence of the consistency of the School AMPS measures, there still remains a need for further studies. A design based on the administration of two complete School AMPS evaluations would have been preferred (Cohen & Swerdlik, 2010). For example, future studies where the students are evaluated with the School AMPS at different points in time, with no intervening intervention, would allow us to evaluate the consistency of the School AMPS measures over time (i.e., estimate test-retest reliability or the coefficient of stability). An even stronger design would be to combine the evaluation of stability with an evaluation of equivalence.

In such a study, the students would perform different School AMPS tasks each time they are evaluated (e.g., Tasks A+B vs. Tasks C+D) to establish a coefficient of stability and equivalence (Cronbach, 1947). An additional consideration for future research might be to evaluate the consistency of the School AMPS measures for different subgroups (e.g., severity of disability; diagnosis, age).

Evidence of validity based on internal structure and consequences of testing

Validation of an instrument is an ongoing process of gathering evidence of validity to support relevant and meaningful interpretation of test scores (Cook & Beckman, 2006; Downing, 2003). Earlier research related the validation of the School AMPS scales (i.e., internal scale validity) had revealed that all of the school motor and school process items, and all of the School AMPS tasks demonstrate acceptable goodness of fit to the MFR model of the School AMPS. An analysis of School AMPS data for students from NA, Europe, and ANZ had revealed (1) no differential task functioning, (2) minimal DIF for Endures, Walks, and Paces; and (3) no DTF. Similarly, earlier analysis of School AMPS data for boys and girls revealed no evidence of differential item or task functioning by gender (Fisher, et al., 2007). In the research that comprised this thesis, we continued the process of gathering evidence to support the validity of the School AMPS scales.

Stability of the skill items (DIF)

More specifically, in our process of evaluating for *evidence of validity based on internal structure*, we evaluated for DIF to determine if the skill items in the School AMPS function differently for different groups of students. Given the potential implications for occupational therapists in Nordic countries, and more specifically, for occupational therapy practice focused on children with mild disabilities within schools in Sweden, we chose to evaluate if the School AMPS skill item hierarchies remain stable among four different world regions (NA, ANZ, UK, and the Nordic countries) (Study II), as well as between typically-developing students and students with mild disabilities (Study III).

Overall, we found minimal evidence of DIF. That is, while the school motor skill items Walks, Moves, and Endures, and the process skill

item Navigates displayed between-region DIF, none of the school motor or school process skill items displayed DIF when we compared each world region to the reference values for the combined item difficulty calibrations of all four world regions. We consider the result from this latter comparison to be of most importance clinically because when the School AMPS computer-scoring program (Three Star Press, 2005) is used to generate the student's quality of schoolwork performance measures, they are based on total sample calibrations. Thus, our results suggest that occupational therapists can use the School AMPS computer-software program to obtain valid student school motor and school process quality of performance measures without concern that minimal between-region DIF disrupts the School AMPS measurement system (Conrad, Dennis, Bezruczko, Funk, & Riley, 2007).

The school motor skill items Walks, Transports, Lifts, and Endures displayed DIF when we compared the relative skill item hierarchies for the children with mild disabilities to those for the typically-developing children, but when we calculated the standardized difference (Z) (Wright & Masters, 1982; Wright & Stone, 1979) we found that the only paired item difficulty calibration values that differed significantly were for Endures. Moreover, when we examined the data for these items further, we found that three of the four school motor items (Walks, Transports, Lifts) were poorly targeted, resulting in high estimation error due to a high rate of maximum raw scores.

The persistent evidence that Endures displays DIF in different contexts may suggest that there exist unconfirmed group differences and/or that a secondary dimension is present in the School AMPS motor scale (Osterlind & Everson, 2009). However, when we examined the data for Endures (Study III), we found that raters had given unusually low scores on Endures to typically-developing students, suggesting that the significant difference in skill item difficulty calibration values was due to rater scoring error. That is, the skill item Endures pertains to *obvious* evidence of physical fatigue during schoolwork task performance, something that is rarely observed when typically-developing students perform their schoolwork tasks. It may be that occupational therapists incorrectly scored these students down on Endures “just because” the students said that they were tired. Yet, there is no logical reason, clinically or

theoretically, to think that typically-developing student should demonstrate more problems performing schoolwork tasks without obvious evidence of physical fatigue than do those with mild disabilities. There also is no logical reason that we were able to identify as to why raters might be more likely to score the typically-developing students lower on Endures than they did the children they tested who had mild disabilities. Finally, even though Endures demonstrated DIF, we believed that this was of minimal risk to the validity of the School AMPS motor scale as a single item likely has little or no effect on the estimation of student's School AMPS quality of schoolwork performance measures.

Our validity evidence based on the internal structure of the School AMPS scales generally supports the clinical use of the School AMPS measures when testing students in Nordic countries. Our results also have implications for conducting international research studies, and for comparing cross-regional school motor and school process measures obtained in clinical practice. Obviously, our results pertain only to the four world regions and two “diagnostic” groups we studied. Further research studies are needed to explore for DIF among additional diagnostic groups, as well as based on other cultural or ethnic grouping criteria (e.g., Asian countries vs. western countries, industrialized vs. developing countries, social class). One focus of such future research might be to determine if Endures (or any other of the skill items that showed DIF) persists in displaying DIF when other groups are compared.

Differential test functioning (DTF)

When we analyzed data for a larger sample, and separated the data for Europe into data for the Nordic countries, data for UK, and data for other European countries, we again found no evidence of DTF among NA, ANZ, UK, and the Nordic countries (Study II). Thus, when we considered our results in relation to validity *evidence related to consequences of testing*, we concluded that the presence of slight DIF was not a threat to the measurement system – there was no evidence of test bias (Camilli & Shepard, 1994). This result adds to already existing evidence that supports cross-regional use of the School AMPS and making valid comparisons of School AMPS results among NA, ANZ, UK, and the Nordic countries. As I discussed above in relation to cross-regional DIF, future research is still needed in

relation to evaluation of test bias among different cultural or ethnic groups.

Identification of actual differences in skill challenge

The results from Study III tentatively provided further evidence of validity related to *consequences of testing* in that we were able to identify individual School AMPS skill items that are actually more or less challenging to perform for students with mild disabilities than for their typically-developing peers in the context of performing schoolwork tasks. Yet, in contrast to our expectations, the school motor skills that were actually harder for the students with mild disabilities to perform were Transports, Lifts, Walks, Moves, Aligns, Paces, and Positions (see Table 9). These are the school motor skill items that generally are the easiest for all students to perform or school motor skill items that are hardest for all students to perform. Thus, they may have emerged, as I noted earlier in relation to relative differences (i.e., DIF), more because of off-targeting than because they have any clinical meaning. Given the difficulties students with sensory integration and/or developmental coordination disorders commonly display related to fine motor coordination and poor handwriting (Bundy, et al., 2002; Missiuna, et al., 2004), we instead would have expected school motor skills such as Calibrates, Manipulates, Grips, and Coordinates to have emerged. That they did not emerge may support our speculation that error due to off-targeting was indeed a factor impacting our results. If so, we cannot conclude that our findings contribute in any clinically meaningful manner to validity evidence for the School AMPS – the school motor skill items that did emerge are not ones that occupational therapists would feel any reason to address in an intervention program.

In contrast, our finding that the majority of the school process skill items were actually more difficult to perform for students with mild disabilities (see Table 10) likely do reflect their clinically meaningful difficulties during occupational performance. These difficulties include temporal organization of task actions, and organizing task space and objects used during the performance of schoolwork tasks. They also demonstrated more actual difficulty in adapting their performance to avoid and overcome problems with task performance.

We had reasoned that the clinical benefit of the results of an analysis of actual differences might lead to the design of effective interventions for children with mild disabilities. Given the questionable meaningfulness of the results of our analysis of the school motor skill items, we hesitate to suggest that our results have any potential to be clinically useful. In contrast, our results for the school process items may be more useful clinically, but future research will be needed to determine if any benefits can be derived. Such research might involve designing a program that progresses from (1) providing interventions focused on the easiest items for all students to perform, but which were actually more difficult for those with mild disabilities (i.e., Chooses, Searches/Locates); to (2) providing interventions that are focused on those school process skills that are hardest for all students to perform (i.e., Benefits, Accommodates). This program might be compared to a program that begins with a focus on those school process skills that were only slightly harder for students with mild disabilities to perform (e.g., Attends Gathers, Chooses) and then gradually progressing to those that were much more challenging for them to perform (e.g., Accommodates, Adjusts, Notices/Responds). In either case, the rationale for the progression would be based on beginning with those school process skills that might be most easily developed through intervention.

Evidence based on relations to other variables

When we planned the studies included in this thesis, there was evidence that the School AMPS quality of performance measures demonstrate expected positive relationships with age, and are sensitive enough to identify differences in quality of schoolwork task performance between typically-developing students and those who have identified developmental/neurological, cognitive/psychological, or other/multiple disabilities (Fisher, et al., 2007). There was also evidence that the School AMPS measures were sensitive enough to identify differences in quality of schoolwork task performance between typically-developing students and those who were at risk (Fisher & Duran, 2004).

We, therefore, decided to gather additional validity *evidence based on relations to other variables* by cross-validating Fisher and Duran's (2004) results, and evaluating if the School AMPS quality of performance measures were also sensitive enough to detect

differences between typically-developing students and those with mild disabilities (Study III). Given that there was preliminary evidence that the two groups (i.e., at risk and mild disabilities) do not demonstrate significant differences in mean school motor or process quality of performance measures (Fisher, et al., 2007), and that the two groups have been described as being “virtually identical” behaviorally (Johnson, 1998), we expected and found significant differences in quality of schoolwork task performance between students with mild disabilities who were between 4 to 11 years of age and their age-matched typically-developing peers. Our results indicated that students with mild disabilities demonstrate increased clumsiness or effort, decreased time and space efficiency, and/or greater need for assistance from others when performing schoolwork tasks in natural classroom environments. It would have been preferred to match students by 1-year age groups, but if we had attempted to do so, we would have had to remove an additional 50 students. We made our decision to retain as many students in each diagnostic group as was reasonably possible so as to ensure stable estimates of the item difficulties for both groups.

Considered together with the results of earlier research, these results support the sensitivity of the School AMPS measures to detect differences in school motor and school process quality of schoolwork task performance between groups of students who are expected to vary. The magnitude of the differences in the overall mean school motor and school process quality of performance between the two groups in Study II (typically-developing vs. mild disabilities) was also reflected in the large effect sizes, supporting the conclusion that the two groups differ. Moreover, applying the criteria of 0.30 logit or greater for a meaningful difference between groups it is likely that these differences were clinically observable (Fisher, et al., 2007).

There is, however, a need for continued research. For example, it will be important to replicate our study and that of Fisher and Duran (2004) with a larger sample that extends over a longer age span (e.g., 3–12 years of age). Such a study could be designed to compare all three groups (typically-developing, at risk, mild disabilities) and evaluate if any identified differences remain stable, increase, or decrease with increasing age.

In Study IV, we gathered additional validity *evidence based on relations to other variables* by evaluating if the School AMPS measures are sensitive enough to be clinically useful for evaluating

change following intervention. We found statistically significant improvements in school motor and school process quality of schoolwork task performance, which provided us with preliminary evidence to support the validity of the School AMPS measures for such purposes.

When viewed from the perspective of occupational therapy services, something that ultimately was of interest to me, our results also represent the first evidence of the effectiveness of school-based occupational therapy services for improving a student's quality of occupational performance when interventions are provided in collaboration with the teacher in a Swedish context. This suggests that not only students with mild disabilities, but also students at risk, who also struggle to perform their schoolwork tasks, may benefit from interventions when provided in the natural school context. It is also interesting to note that even though 20 students simultaneously received special educational services, these services had no effect on the student's quality of schoolwork task performance. The special educators' focus is academic, not quality of schoolwork task performance. The occupational therapists' contributions to the team compliment, not replace, the important contributions of the special educator.

This study is also of clinical relevance because of the current emphasis on integrating related services into the natural classroom environment versus performing "pull-out" therapy (Hinder & Ashburner, 2010). The former mode for providing services is based on an educational model, where the focus is on ensuring a match between the student's abilities and the challenges of the task he or she is to perform. Such interventions are based on consulting with teacher, and providing needed services directed toward enabling the teacher to modify his or her teaching strategies (e.g., building in more structure in classroom assignments or activities) and/or the classroom environment (e.g., diminishing noise, modifying the types of supports the student is given). This mode of service is in contrast to services based on a bottom-up, medical model where the focus is on identifying the student's underlying impairments (e.g., poor fine motor coordination, distractibility), and then providing interventions focused directly on trying to remediate those impairments. Through the use of a collaborative working relationship with the student's teacher, it was possible to fully integrate the interventions into the classroom.

This strategy enabled the students to remain in the classroom with their peers, and not feel stigmatized or that they “missed out” on what was going on. Thus, it might be important to share with school administrators our findings that occupational therapy provided in the classroom did have an impact on the majority of the children in our sample, suggesting that students at risk and students with mild disabilities may benefit from occupational therapy services that are integrated into the classroom.

Paulsen (2008) has stated that effective school-based collaboration must be sustained over time if successful programs are to be developed and implemented. There are, however, many important factors that must be inherent in the collaborative process. Such factors may relate to effective communication skills among all team members who then are able to draw upon each others’ expertise (Hinder & Ashburner, 2010). That is, one person cannot be expected to be an expert in everything. Moreover, the school principal must prioritize and provide time for teachers to engage in collaborative activities as well as support the idea of collaboration among all persons involved (Paulsen, 2008; Vincent, Stewart, & Harrison, 2008). Finally, time efficiency such as well planned meetings is of utmost importance (Paulsen, 2008). That is, even though teachers perceive consultative services to be effective (Dreiling & Bundy, 2003; Dunn, 1990), one of the most common barriers for successful collaboration is lack of time available (Reid, Chiu, Sinclair, Wehrmann, & Naseer, 2006; Vincent, et al., 2008).

The demonstration project described in this thesis was new and, therefore, not well established. Not all school principals supported the project which prevented occupational therapy services from being implemented in some schools. Moreover, among those students who were not tested with the School AMPS twice (see Figure 2), organizational- and time-related issues were a major reason for “breakdown” in the consultative process. We have no way of knowing what our results would have been had we been able to prevent or overcome these obstacles.

Nevertheless, this study provides *preliminary* pilot evidence that supports the involvement of occupational therapists who implement occupation-based assessment and consultative adaptive interventions to students at risk or with mild disabilities who currently do not receive such services. Thus, it seems that

occupational therapists likely have an important role in schools to support students at risk or with mild disabilities. Preferably *all* students in need should have access to related services as stated in laws and national agreements in Sweden and in many other countries (e.g., United States, Australia, New Zealand, Canada, the UK) (IDEA, 2004; SFS, 1995; UNESCO, 1994; United Nation, 2002). The Salamanca Statement (UNESCO, 1994) states that occupational therapists are one resource who can play a lead role in supporting the educational needs of students. This may be an indication for change in practices in Sweden – to include occupational therapy as a related service in schools where it is not already included. This could be done by implementing *best practice* and moving toward a vision of providing client-centered and occupation-based services in early intervention and school-based programs (AOTA, 2004; Dunn, 2000; Simmons Carlsson, Hocking, & Wright-St Clair, 2007).

Obviously, several limitations are inherent in the design of our pilot intervention study and we are aware that the improvements we observed cannot confidently ascribed to the occupational therapy interventions. For example, we only had a small control group. Thus, it is not possible to infer causality – that the interventions alone were responsible for the improvements in quality of schoolwork task performance. Additionally, all available data were retrieved retrospectively from the demonstration project related to implementing school-based occupational therapy into the school system, and the demonstration project was not designed to be a research project. Thus, there was no pre-planned control group, nor was there any attempt to control time between evaluations. The time factor, however, was not considered a major threat to internal validity of the design of this study as a recent study has revealed that the yearly mean increase in quality of schoolwork task performance among students at risk 5 to 10 years of age rarely exceeds 0.10 logit ($M \leq 0.03$ logit) (Sperens, et al., 2010). Of greater potential threat was the fact that the same occupational therapist who implemented the School AMPS evaluations also planned and in part implemented the interventions (i.e., consulted with the teacher who in turn implemented compensatory strategies) and performed the follow-up evaluations. The occupational therapists, however, were essentially blind to the final school motor and school process quality of schoolwork performance measures at time of scoring the second School AMPS evaluation. That is, the intervention phase was several months in length and the majority of the students also

performed different tasks at reevaluation. The time interval likely would make it very hard for the occupational therapist to remember how the student was scored on the initial evaluation. Moreover, because the students performed different tasks, the occupational therapist would have no way of knowing the students' quality of schoolwork task performance measures until after the data have been analyzed using the School AMPS computer-scoring program (Three Star Press, 2005). Nevertheless, we cannot be sure that the occupational therapists' desire to see changes did not impact our results.

It is evident that further research is needed in this area. In existing studies, sample sizes have been small, designs have been weak, and interventions and outcomes have mainly focused on functional impairments of children with mild disabilities (fine-motor coordination, cognition, perception) (Chu & Reynolds, 2007; Taylor, Fayed, & Mandich, 2007; Vandenberg, 2001). Obviously, there is the need to replicate our study, ideally using a randomized control trial design (Hohmann & Shear, 2002). We also recommend the use of a variety of outcome measures, including goal attainment (Mailloux, et al., 2007), as well as evaluation of perceived participation by students', teachers, and parents. We also feel it is especially important to involve the student in the process of planning goals and interventions to facilitate measurement of outcomes that are of significance for the individual student.

Finally, when we used different criteria to evaluate for changes following intervention, we obtained different results. This clearly underscores that the criterion used to evaluate the results of an intervention is of utmost importance, and must be chosen carefully. For example, if it is judged to be "enough" evidence when a student makes gains that are clinically observable, then our results suggest that 80.5% of the students demonstrated improvements in school motor and/or school process quality of occupational performance. While these differences do have clinical meaning in that they are observable, they are not necessarily "scientific" if the traditional p value of .05 is used as the criteria for statistical significance. Moreover, if one considers statistical significance to be the desired criterion, we will obtain different results depending on the p value we use to evaluate for significance. That is, when we used Harvill's (1991) guideline for statistical significance, based on a change greater than 2 SEm ($p < .15$), we found that the number of students

who showed improvements in school motor and/or school process quality of schoolwork performance decreased from 80.5% to 66.7%. Had we used standardized Z and set our criterion at $p < .05$ the number showing improvements would have dropped to only 27.8% of our sample.

From a clinical perspective, when implementing a program evaluation, it may be enough to gather evidence for clinically meaningful changes. Such evidence of change might be supplemented by gathering data related to percentage of student goals achieved. When implementing outcomes research, we recommend the use of Harvill's (1991) strategy. We base this conclusion on our finding of both significant differences in mean school motor and school process quality of performance measures and large effect sizes when the group as a whole was considered, not just the number of students who showed meaningful improvements.

Other methodological considerations

We used Pearson product-moment correlations (r) as our index of reliability. However, this index has been criticized as it only determines the strength of linear association between two variables, and does not measure agreement. Thus, the major disadvantage of using Pearson r is that two measures can have a perfect correlation, but yet differ systematically. Therefore, the use of intra-class correlation methods is recommended to overcome this problem (Domholdt, 2000). While, we agree with this perspective, we felt that we had minimal risk of a systematic difference given that we had equivalent means, standard deviations, and standard errors of measurement between our different split-halves.

The method used to evaluate DIF in Study III was based on an older method than was used in Study II, and we also used different criteria for determining the size of DIF. In Study III, we used ≥ 0.43 logit based on (Bernspång & Fisher, 1995; Cooke, et al., 2000; Duran & Fisher, 1996; Goldman & Fisher, 1997; Magalhães, et al., 1996; Oakley, et al., 2003; Silverstein, Fisher, et al., 1992; Stauffer, Fisher, & Duran, 2000) and in Study III we used ± 0.55 logit based on Tristán (2006). Standards for important effect sizes, however, are lacking,

but recommended values typically range from 0.40 logit to 0.60 logit (Conrad & Smith, 2004; Draba, 1977; Linacre, 1994; Tristán, 2006). The most strict values were developed by Paek (personal communication, January 7, 2009), who transformed Rasch-modelled logit differences into effect sizes based on the Educational Testing Services (ETS) method for classifying DIF (Wilson, 2005). Paek found that a logit difference of <0.426 is negligible, 0.426 to 0.638 is slight to moderate, and >0.638 is moderate to severe. It remains common practice, however, to consider values <0.50 logit as evidence of no DIF (Draba, 1977; Tennant & Pallant, 2007; Tristán, 2006), because “measures based on item calibration[s] with random deviations up to 0.50 logit are ‘for all practical purposes free from bias’” (Linacre, 1994, p. 328). Conrad et al. (2007) used a more liberal 0.60 logit criterion, based on Norman, Sloan, and Norwich (2003), who recommended the use of half of a standard deviation as an indicator of a clinically important difference. Perhaps the most comprehensive analysis to determine a critical value for an important effect size was implemented by Tristán. He developed a method for evaluating for significant DIF based on the use of normalized *SEms*, where $SEm_{\text{normalized}} = (SEm \text{ (SqRt of } N))/10$. Tristán found that when *SEms* were normalized, the minimum possible *SEm* is 0.20 logit. With *SEm* values of 0.20 logit, a difference in item calibration values of 0.55 logit would be required for statistical significance. That is, the statistical test for a significant difference in item calibration values is: $t = D_1 - D_2 / \text{SqRt}(SEm_1^2 + SEm_2^2) = 1.96$ (Wright & Stone, 1979). Thus, $0.55 \text{ logit calibration difference} / \text{SqRt}(0.20^2 + 0.20^2) = 1.96$; $p = 0.05$. We chose, therefore, to set our criteria for the presence of significant DIF based on a logit difference of at least ± 0.55 logit between world regions. Thus, when the skill item hierarchies remain stable (i.e., the difference between world regions is less than ± 0.55 logit), there is no statistical DIF.

Summary of clinical implications

The findings of this thesis have several clinical implications concerning the evaluation of occupational performance in school settings using the School AMPS as the outcome measure. The results also provide important information about the potential benefits of the role of occupational therapy when implemented through collaborative

consultation with the teacher. The following are a summary of the clinical implications of the studies included in this thesis:

- The consistency of reliability evidence for the School AMPS measures supports valid interpretations of students' school motor and school process quality of performance measures.
- Occupational therapists using the School AMPS in the process of making decisions about individual students, planning occupation-based interventions, and performing follow-up evaluations to measure outcomes may compare the quality of schoolwork performance measures directly by calculating the magnitude of difference to identify when statistically significant differences likely have occurred.
- The School AMPS can be used as a valid assessment across four major world regions enabling comparisons of quality of schoolwork performance measures between students from these four world regions.
- The School AMPS can be used to evaluate the consequences of mild disabilities on the students' quality of schoolwork task performance.
- The identification of the hierarchy of actual differences in school process skills between typically-developing students and students with mild disabilities may have potential for informing the planning of a future intervention study.
- The School AMPS appears to be a sensitive measure that can be used to evaluate the effectiveness of interventions for students at risk or with mild disabilities.
- When provided with adequate support or adaptations, students at risk or with mild disabilities can perform their schoolwork tasks with less clumsiness, more efficiency, and/or with less assistance from others following consultative occupational therapy interventions provided within the general education classroom.

Conclusions

The conclusions of Studies I to IV will be addressed in accordance with the overall focus of this thesis, to further develop and validate the School AMPS, and the secondary focus, to contribute to knowledge about children with mild disabilities and the usefulness of the School AMPS for planning and implementing school-based occupational therapy services.

- Cross-validation using available methods to evaluate consistency of the School AMPS measures revealed high reliability coefficient estimates for both school motor and school process quality of schoolwork performance measures, irrespective of which schoolwork tasks the included students had performed.
- The overall *SEm* values for the School AMPS scales provide reasonable approximations for use to compare students school motor and school process quality of schoolwork performance measures over time.
- The overall stable skill item hierarchies among four major world regions indicate evidence of validity based on internal structure of the School AMPS. Thus, the School AMPS can be used to provide valid quality of schoolwork performance measures at the group level among students from these different world regions.
- Overall, the relative skill item hierarchies remained stable between typically-developing students and students with mild disabilities, also indicating evidence of validity based on internal structure of the School AMPS. Students with mild disabilities demonstrated more limited ability to perform the majority of the School AMPS skill items compared to their typically-developing peers, but they were also equally skilled in performing some skill items as are typically-developing students.
- When occupational therapy services are implemented using a top-down, occupation-based consultative approach, the School AMPS quality of schoolwork performance measures can be used to detect changes in quality of schoolwork task performance among students at risk or with mild disabilities.

- While the studies in this thesis have generated evidence of reliability and validity of the School AMPS measures, there still remains a need for future research to continue the process of gathering evidence that support the use of the School AMPS in the process of evaluating students occupational performance needs in school settings.

Acknowledgements

What an adventure this has been! I entered into this journey full of passion and with a dream about implementing school-based occupational therapy in the school system. Now I have ended this journey with a PhD degree and I will continue to advocate for those children who “fall between the cracks” in school. It has not been easy, but I have really enjoyed being a doctoral student. This adventure has been full of challenges and at times extreme frustration along with a good amount of joy. I could not have done this alone and I wish to express my deepest gratitude to all of you who have supported me and contributed in different ways to enabling me to complete this work.

I would especially like to thank all teachers and children who have contributed with their time during data collection and for sharing your experiences in your everyday life at school. Thank you also to *Ingela Gotthardsson* for your endless support and believing in implementing school-based occupational therapy. I have learned a lot and you made my dream come true.

I wish to express my warmest gratitude to my main supervisor *Anne Fisher* for inviting me onboard on this adventure and into your life, for your never ending engagement and for believing in me all the way. I cannot thank you enough for all the thousands of hours you have spent reading and editing my manuscripts and your patience when my brain was dead, and for calming me down during states of frustration and struggles to understand. I also wish to thank you for sharing your invaluable knowledge about children and occupational therapy, for guiding me through client-centeredness, Rasch measurement, skill item hierarchies, DIF, standard errors of measurement, and much more, and you were right: “It passed”!

Britta Löfgren, co-supervisor and co-author, thank you for taking an interest in my studies, for your professional advice during different stages of the process, and for your warm support and belief in my abilities.

Brett Berg, co-author, thank you for the countless of hours you have spent running Rasch analyses, for always being on “stand by” by

Acknowledgements

your computer, providing me with new data files, references, valuable ideas and opinions, and for sending me fun and encouraging e-mails.

Lou Ann Griswold, thank you for sharing your valuable knowledge about school-based occupational therapy, and for providing me with Tang, muffin forms, and sandwich bags.

Birgitta Bernspång, thank you for your generosity, warm support, and encouragement to move on, and for believing in the importance of occupational therapy in school settings.

Ingeborg Nilsson, my dear colleague, thank you for your guidance when I first became a doctoral student, for your energy, encouragement, and support during all phases of this adventure. Thank you also for being my friend, but where were you during these last days when I needed you, and who will help with making the 2010 gingerbread house? I will miss you and your wonderful family on the day of my defense.

Ulrica Cavallin, my colleague and friend, thank you for your energy and interesting and inspiring discussions. Where do we go from here?

Anders Kottorp, thank you so much for taking your time to read the final draft of my thesis and for preparing challenging questions.

Special thanks to all my colleagues within the doctoral group who have provided ongoing support and friendship during my years as a doctoral student. Thank you also for valuable discussions and feedback particularly during the final phase of this work. Good luck to you all!

All my colleagues at the Department of Community Medicine and Rehabilitation, Division of Occupational Therapy at Umeå University, thank you for always being encouraging and supportive. Special thanks to *Bodil Nilsson* and *Gunilla Ottoson* for believing in my dream. You were the ones who first made me interested in research. Special thanks also to *Margareta Lindberg* who encouraged me in the first place to begin this adventure, and to *Bertil Sawert* for always reminding me that there is more in life than work and for asking me to “breathe”.

Thank you also to all our friends for opportunities to relax, and share dinners, fun, and joy. Special thanks to *Anna* and *Danne* for all your support, and for taking *Elin* and *Ebba* into your home and for feeding me and my family from time to time. Also to *Tilda*, *Maja*, and *Ture*, for all your energy and entertainment – you are priceless!

Finally, but most of all, my deepest gratitude to my beloved husband *Michael* for your stunning patience and endless love, and to my precious children *Casper*, *Elin*, and *Ebba* – you are best and I am so proud of you!

The studies presented in this thesis would have been impossible to perform without the generous financial support of the *Swedish Research Council (Vetenskapsrådet)* and the *Medical Faculty at Umeå University*, Sweden.

References

- Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *The standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Occupational Therapy Association [AOTA] (2004). Occupational therapy services in early intervention and school-based programs. *American Journal of Occupational Therapy*, *59*, 681-685.
- American Occupational Therapy Association [AOTA] (2008). Occupational therapy practice framework: Domain and process (2nd ed.). *American Journal of Occupational Therapy*, *62*, 625-683.
- American Occupational Therapy Association, A. (2002). Occupational therapy practice framework: Domain and process. *American Journal of Occupational Therapy*, *56*, 609-639.
- American Psychiatric Association [APA] (2004). *Diagnostic and statistical manual of mental disorders* (4th ed.). Arlington, VA: Author.
- Ary, D., Jacobs, L. C., & Razavieh, A. (2002). *Introduction to research in education* (6th ed.). Belmont, CA: Wadsworth Group.
- Atchison, B. T., Fisher, A. G., & Bryze, K. (1998). Rater reliability and internal scale and person response validity of the School Assessment of Motor and Process Skills. *American Journal of Occupational Therapy*, *52*, 843-850.
- Barnes, K. J., & Turner, K. D. (2001). Team collaborative practices between teachers and occupational therapists. *American Journal of Occupational Therapy*, *55*, 83-89.
- Bayona, C. L., McDougall, J., Tucker, M. A., Nichols, M., & Mandich, A. (2006). School-based occupational therapy for children

References

- with fine motor difficulties: Evaluating functional outcomes and fidelity services. *Physical & Occupational Therapy in Pediatrics*, 26, 89-110.
- Becker, G. (2000). How important is transient error in estimating reliability? Going beyond simulation studies. *Psychological Methods*, 5, 370-379.
- Bernspång, B., & Fisher, A. G. (1995a). Differences between persons with right or left cerebral vascular accident on the Assessment of Motor and Process Skills. *Archives of Physical Medicine and Rehabilitation*, 76, 1144-1151.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Boomsma, A., van Duijn, M. A. J., & Snijders, T. A. B. (Eds.). (2001). *Essays on item response theory*. New York: Springer-Verlag.
- Bose, P., & Hinojosa, J. (2008). Reported experiences from occupational therapists interacting with teachers in inclusive early childhood classrooms. *American Journal of Occupational Therapy*, 62, 289-297.
- Brown, T., & Chien, C. W. (Eds.). (2010). *Occupation-centred assessment with children*. Oxford: Wiley-Blackwell.
- Bundy, A. C. (1995). Assessment and intervention in school-based practice: Answering questions and minimizing discrepancies. *Physical & Occupational Therapy in Pediatrics*, 15, 69-88.
- Bundy, A. C., Lane, S. J., & Murray, E. A. (2002). *Sensory integration theory and practice* (2nd ed.). Philadelphia: F. A. Davis.
- Burtner, P. A., McMains, M. P., & Crowe, T. K. (2002). Survey of occupational therapy practitioners in southwestern schools: Assessments used and preparation of students for school-based practice. *Physical & Occupational Therapy in Pediatrics*, 22, 25-39.
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221-256). Westport, CT: Praeger.

- Camilli, G., & Shepard, L. A. (1994). *MMSS methods for identifying biased test items* (Vol. 4). Thousand Oaks, CA: Sage Publications.
- Canadian Association of Occupational Therapists [CAOT] (1997). *Enabling occupation: An occupational therapy perspective*. Ottawa, ON: CAOT Publications ACE.
- Cantell, M. H., Smyth, M. M., & Ahonen, T. P. (1994). Clumsiness in adolescence: Emotional, motor, and social outcomes of motor delay detected at 5 years. *Adapted Physical Activity Quarterly*, *11*, 115-129.
- Case-Smith, J., & Cable, J. (1996). Perceptions of occupational therapy regarding service delivery models in school-based practice. *Occupational Therapy Journal of Research*, *16*, 23-44.
- Case-Smith, J., Law, M., Missiuna, C., Pollock, N., & Stewart, D. (2010). Foundations for occupational therapy practice with children. In J. Case-Smith (Ed.), *Occupational therapy for children* (6th ed.). Maryland Heights, MO: Mosby Elsevier.
- Case-Smith, J., & O'Brien, J. C. (2010). *Occupational therapy for children* (6th ed.). Maryland Heights, MO: Mosby Elsevier.
- Chen, H. F., & Cohn, E. S. (2003). Social participation for children with developmental coordination disorder: Conceptual, evaluation, and intervention considerations. *Physical & Occupational Therapy in Pediatrics*, *23*, 61-78.
- Chu, S., & Reynolds, F. (2007). Occupational therapy for children with attention deficit hyperactivity disorder (ADHD), part 1: A model of practice. *British Journal of Occupational Therapy*, *70*, 372-383.
- Cohen, J. (1988). *Statistical power analyses for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, R. J., & Swerdlik, M. E. (2010). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). New York: McGraw-Hill.

References

- Conrad, K. J., Dennis, M. L., Bezruczko, N., Funk, R. R., & Riley, B. B. (2007). Substance use disorder symptoms: Evidence of differential item functioning by age. *Journal of Applied Measurement, 8*, 373-387.
- Conrad, K. J., & Smith, E. V. (2004). International conference on objective measurement: Applications of Rasch analysis in health care. *Medical Care, 42*, 1-6.
- Cook, B. G. (2001). A comparison of teachers' attitudes toward their included students with mild and severe disabilities. *Journal of Special Education, 34*, 203-213.
- Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *American Journal of Medicine, 119*, 7-16.
- Cooke, K. Z., Fisher, A. G., Mayberry, W. L., & Oakley, F. (2000). Differences in ADL process skills of persons with and without Alzheimer's disease. *Occupational Therapy Journal of Research, 20*, 87-104.
- Coster, W. (1998). Occupation-centered assessment of children. *American Journal of Occupational Therapy, 52*, 337-344.
- Coster, W., Deeney, T., Haltiwanger, J., & Haley, S. (1998). *School Function Assessment*. San Antonio, TX: Psychological Corporation.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart, & Winston.
- Cronbach, L. J. (1947). Test "reliability": Its meaning and determination. *Psychometrika, 12*, 1-16.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- Cunningham, C. E., & Boyle, M. H. (2002). Preschoolers at risk for attention-deficit hyperactivity disorder and oppositional defiant disorder: Family, parenting, and behavioral correlates. *Journal of Abnormal Child Psychology, 30*, 555-569.
- Dawson, B., & Trapp, R. G. (2004). *Basic & clinical biostatistics* (4 ed.): New York: Appleton & Lange.

- Dewey, D., Kaplan, B. J., Crawford, S. G., & Wilson, B. N. (2002). Developmental coordination disorder: Associated problems in attention, learning, and psychosocial adjustment. *Human Movement Science, 21*, 905-918.
- Dewey, D., & Wilson, B. N. (2001). Developmental coordination disorder: What is it? In C. Missiuna (Ed.), *Children with developmental coordination disorder: Strategies for success* (pp. 5-27): Binghamton, NY: Haworth Press.
- Domholdt, E. (2000). *Rehabilitation research: Principles and applications* (3th ed.). St. Louis, MO: Elsevier Saunders.
- Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education, 37*, 830-837.
- Draba, R. E. (1977). *The identification and interpretation of item bias*. Chicago: MESA Press.
- Dreiling, D. S., & Bundy, A. C. (2003). A comparison of consultative model and direct-indirect intervention with preschoolers. *American Journal of Occupational Therapy, 57*, 566-569.
- Dunbar, S. B. (2007). *Occupational therapy models for intervention with children and families*. Thorofare, NJ: SLACK.
- Dunn, W. (1990). A comparison of service provision models in school based occupational therapy services: A pilot study. *Occupational Therapy Journal of Research, 10*, 300-320.
- Dunn, W., Brown, C., & McGuigan (1994). The ecology of human performance: A framework for considering the effect of context. *American Journal of Occupational Therapy, 48*, 595-607.
- Dunn, W. (2000). *Best practice occupational therapy: In community service with children and families*. Thorofare, NJ: SLACK.
- Duran, J. L., & Fisher, A. G. (1996). Male and female performance on the Assessment of Motor and Process Skills. *Archives Physical Medicine and Rehabilitation, 77*, 1019-1024.
- Dussart, G. (1994). Identifying the clumsy child in school: An exploratory study. *British Journal of Special Education, 21*, 81-86.

References

- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3th ed.). New York: Macmillan Publishing.
- Fisher, A. G. (1993). The assessment of IADL motor skills: An application of many-faceted Rasch analysis. *American Journal of Occupational Therapy*, *47*, 319-329.
- Fisher, A. G. (2006). Performance skills: Definitions and evaluation in the context of the Occupational Therapy Practice Framework. In H. Pendleton & W. Schultz-Krohn (Eds.), *Pedretti's occupational therapy: Practice skills for physical dysfunction* (6th ed., pp. 372-402). St. Louis, MO: Mosby.
- Fisher, A. G. (2009). *Occupational therapy intervention process model: A model for planning and implementing top-down, client-centered, and occupation-based occupational therapy interventions*. Fort Collins, CO: Three Star Press.
- Fisher, A. G., Bryze, K., & Atchison, B. T. (2000). Naturalistic assessment of functional performance in school-settings: Reliability and validity of the School AMPS scales. *Journal of Outcome Measurement*, *4*, 491-512.
- Fisher, A. G., Bryze, K., Hume, V., & Griswold, L. A. (2007). *School AMPS: School Version of the Assessment of Motor and Process Skills* (2nd ed.). Fort Collins, CO: Three Star Press.
- Fisher, A. G., & Duran, G. A. (2004). Schoolwork task performance of students at risk of delays. *Scandinavian Journal of Occupational Therapy*, *11*, 191-198.
- Fisher, A. G., & Jones, K. B. (2010). *Assessment of Motor and Process Skills: Volume 1 - Development, standardization, and administration manual* (7th ed.). Fort Collins, CO: Three Star Press.
- Fox, A. M., & Lent, B. (1996). Clumsy children: Primer on developmental coordination disorder. *Canadian Family Physician*, *42*, 1065-1071.
- Goldman, S. L., & Fisher, A. G. (1997). Cross-cultural validation of the Assessment of Motor and Process Skills (AMPS). *British Journal of Occupational Therapy*, *60*, 77-85.

- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.). Westport, CT: Praeger Publishers.
- Harvill, L. M. (1991). An NCME instructional module on standard error of measurement. *Educational Measurement: Issues and Practice*, 10, 33-41.
- Hellgren, L., Gillberg, C., Gillberg, I. C., & Enerskog, I. (1993). Children with deficits in attention, motor control and perception (DAMP) almost grown up: General health at 16 years. *Developmental Medicine & Child Neurology*, 35, 881-892.
- Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, 34, 177-189.
- Heward, W. L. (2000). *Exceptional children: An introduction to special education* (6th ed.): Upper Saddle River, NJ: Merrill.
- Hinder, E. A., & Ashburner, J. (2010). Occupation-centred intervention in the school setting. In S. Rodger (Ed.), *Occupation-centred practice with children: A practical guide for occupational therapists*. Oxford: Wiley-Blackwell.
- Hocking, C. (2001). Implementing occupation-based assessment. *American Journal of Occupational Therapy*, 55, 463-469.
- Hohmann, A. A., & Shear, M. K. (2002). Community-based intervention research: Coping with the "noise" of real life in study design. *American Journal of Psychiatry*, 159, 201-207.
- Individuals with Disabilities Education Act [IDEA] (2004). *Public Law 108-446 20 U.S.C §1400*.
- Johnson, G. M. (1998). Students at risk: Toward a new paradigm of mild educational disabilities. *School Psychology International*, 19, 221-237.
- Kadesjö, B., & Gillberg, C. (1998). Attention deficits and clumsiness in Swedish 7-year-old children. *Developmental Medicine & Child Neurology*, 40, 796-804.

References

- Kaplan, B. J., Wilson, B. N., Dewey, D., & Crawford, S. G. (1998). DCD may not be a discrete disorder. *Human Movement Science, 17*, 471-490.
- Kielhofner, G. (2008). *Model of Human Occupation* (4 ed.). Baltimore: Lippincott Williams & Wilkins.
- Kopp, S. (2010). *Girls with social and/or attention impairments*. Gothenburg: University of Gothenburg.
- Kottorp, A., Bernspång, B., & Fisher, A. G. (2003). Activities of daily living in persons with intellectual disability: Strengths and limitations in specific motor and process skills. *Australian Occupational Therapy Journal, 50*, 195-204.
- Kramer, P., & Hinojosa, J. (Eds.). (2010). *Frames of reference for pediatric occupational therapy* (3th ed.). Baltimore: Lippincott, Williams & Wilkins.
- Law, M., Baum, C., & Dunn, W. (2001). *Measuring occupational performance: Supporting best practice in occupational therapy*. Thorofare, NJ: SLACK.
- Law, M., Baum, C., & Dunn, W. (2005). *Measuring occupational performance: Supporting best practice in occupational therapy* (2nd ed.). Thorofare, NJ: SLACK
- Law, M., Cooper, B., Strong, B., Stewart, D., Rigby, P., & Letts, L. (1996). The person-environment occupation model: A transactive approach to occupational performance. *Canadian Journal of Occupational Therapy, 63*, 9-23.
- Linacre, J. M. (1993). *Many-faceted Rasch measurement* (2nd ed.). Chicago: MESA Press.
- Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7*, 328.
- Linacre, J. M. (2007). Standard errors and reliabilities: Rasch and raw score. *Rasch Measurement Transactions, 20*, 1086.
- Linacre, J. M. (2009). FACETS: Rasch measurement computer program (Version 3.65.0). Chicago: Winsteps.com.

- Linacre, J. M. (2010). A user's guide to FACETS Rasch-model computer programs (Program manual 3.67.0). Retrieved 2010-11-01 from <http://www.winsteps.com/a/facets.pdf>:
- Linacre, J. M., & Wright, B. D. (2002). Construction of measures from many-facet data. *Journal of Applied Measurement, 3*, 484-509.
- Losse, A., Henderson, S. E., Elliman, D., Hall, D., Knight, E., & Jongmans, M. (1991). Clumsiness in children: Do they grow out of it? A 10-year follow-up study. *Developmental Medicine & Child Neurology, 33*, 55-68.
- Magalhães, L. C., Fisher, A. G., Bernspång, B., & Linacre, J. M. (1996). Cross-cultural assessment of functional ability. *Occupational Therapy Journal of Research, 16*, 45-63.
- Mailloux, Z., May-Benson, T. A., Summers, C. A., Miller, L. J., Brett-Green, B., Burke, J. P., et al. (2007). Goal attainment scaling as a measure of meaningful outcomes for children with sensory integration disorders. *American Journal of Occupational Therapy, 61*, 254-259.
- Mandich, A. D., Polatajko, H. J., Macnab, J. J., & Miller, L. T. (2001). Treatment of children with developmental coordination disorder: What is the evidence? *Physical & Occupational Therapy in Pediatrics, 20*, 51-68.
- Missiuna, C., Rivard, L., & Pollock, N. (2004). They're bright but can't write: Developmental coordination disorder in school aged children. *TEACHING Exceptional Children Plus*. Retrieved 2010-11-20 from http://www.canchild.ca/en/ourresearch/resources/TEC_DCD.pdf.
- Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care, 41*, 582-592.
- Oakley, F., Duran, L., Fisher, A. G., & Merritt, B. (2003). Differences in activities of daily living motor skills of persons with and without Alzheimer's disease. *Australian Occupational Therapy Journal, 50*, 72-78.

References

- Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Oxford English Dictionary (1989). (2 ed.). Oxford: Clarendon Press.
- Pae, T., & Park, G. P. (2006). Examining the relationship between differential item functioning and differential test functioning. *Language Testing, 23*, 475-496.
- Paulsen, K. J. (2008). School-based collaboration. *Intervention in School and Clinic, 43*, 313-315.
- Pennfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In C. R. R. S. Sinharay (Ed.), *Handbook of statistics: Psychometrics* (Vol. 26, pp. 125-167). New York: Elsevier.
- Perrone, M. (2006). Differential item functioning and item bias: Critical considerations in test fairness. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 6*, 1-3
- Polatajko, H. J., & Cantin, N. (2010). Exploring the effectiveness of occupational therapy interventions, other than the sensory integration approach, with children and adolescents experiencing difficulty processing and integrating sensory information. *American Journal of Occupational Therapy, 64*, 415-429.
- Rasmussen, P., & Gillberg, C. (2000). Natural outcome of ADHD with developmental coordination disorder at age 22 years: A controlled longitudinal, community-based study. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 1424-1431.
- Reed, K. L., & Sanderson, S. N. (1999). *Concepts of occupational therapy* (4th ed.). Baltimore: Lippincott, Williams & Wilkins.
- Reid, D., Chiu, T., Sinclair, G., Wehrmann, S., & Naseer, Z. (2006). Outcomes of an occupational therapy school-based consultation service for students with fine motor difficulties. *Canadian Journal of Occupational Therapy, 73*, 215-224.

- Rodger, S. (2010). *Occupation-centered practice with children: A practical guide for occupational therapists*. Malden, MA: Wiley-Blackwell.
- Rodger, S., Brown, G. T., & Brown, A. (2005). Profile of paediatric occupational therapy practice in Australia. *Australian Occupational Therapy Journal*, 52, 311-325.
- Rodger, S., & Ziviani, J. (Eds.). (2006). *Occupational therapy with children: Understanding children's occupations and enabling participation*. Oxford: Blackwell Publishing.
- Sangster, C., Beninger, C., Polatajko, H. J., & Mandich, A. (2005). Cognitive strategy generation in children with developmental coordination disorder. *Canadian Journal of Occupational Therapy*, 72, 67-77.
- Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*, 8, 206-224.
- Schneck, C. M., & Amundson, S. J. (2010). Prewriting and handwriting skills. In J. Case-Smith (Ed.), *Occupational therapy for children* (6th ed.). Maryland Heights, MO: Mosby Elsevier.
- Schoemaker, M. M., & Kalverboer, A. F. (1994). Social and affective problems of children who are clumsy: How early do they begin? *Adapted Physical Activity Quarterly*, 11, 130-140.
- Schumacker, R. E., & Smith, E. V. (2010). Reliability: A Rasch perspective. *Educational and Psychological Measurement*, 67, 394-409.
- Silverstein, B., Fisher, W. P., Kilgore, K. M., & Harley, P. F. (1992). Applying psychometric criteria to functional assessment in medical rehabilitation: II. Defining interval measures. *Archives of Physical Medicine and Rehabilitation*, 73, 507-518.
- Simmons Carlsson, C., Hocking, C., & Wright-St Clair, V. (2007). The "why" of who we are: Exploring the "culture of practice" of Ministry of Education, Special Education occupational therapists and physiotherapists. *Kairaranga*, 8, 6-14.

References

- Skinner, R. A., & Piek, J. P. (2001). Psychosocial implications of poor motor coordination in children and adolescents. *Human Movement Science, 20*, 73-94.
- Skolverket (2001a). *Utan fullständiga betyg: Varför når inte alla elever målen? [Without complete grades: Why do not all students reach the goal?]*. (Report No. 202). Stockholm: Liber Distribution, Publikationstjänst.
- Skolverket (2001b). *Tre magiska G:n: Skolans insatser för elever med funktionshinder [Three magical G's: The school's interventions for students with disabilities]*. (Follow-up study to Report No. 202). Stockholm: Liber Distribution, Publikationstjänst.
- Socialstyrelsen (2003). *Tillgång till habilitering och rehabilitering för barn och ungdomar med funktionshinder: En kartläggning [Access to habilitation and rehabilitation for children and youth with disabilities: A survey]*. (Artikelnr: 2003-103-2). Stockholm, Markör.
- Spencer, K. C., Turkett, A., Vaughan, R., & Koenig, S. (2006). School-based practice patterns: A survey of occupational therapists in Colorado. *American Journal of Occupational Therapy, 60*, 81-91.
- Sperens, M., Munkholm, M., & Fisher, A. G. (2010). Gender differences in quality of schoolwork task performance among typically-developing students and students at risk for or with mild disabilities. *Scandinavian Journal of Occupational Therapy, Early online*, 1-8. doi:10.3109/11038128.2010.533189
- Stauffer, L. M., Fisher, A. G., & Duran, L. (2000). ADL performance of black and white Americans on the Assessment of Motor and Process Skills. *American Journal of Occupational Therapy, 54*, 607-613.
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*, 99-103.
- Svensk Författningssamling [SFS] (1995). *SFS 1995:1100 Skollagen [The Swedish education act]*. Stockholm: Fritzes.

- Swinth, Y., Chandler, B., Hanft, B., Jackson, L., & Shepherd, J. (2003). *Personnel issues in school-based occupational therapy: Supply and demand, preparation, certification, and licensure*. (COPSSE Document No. IB-1). Gainesville, FL: University of Florida, Center on Personnel Studies in Special Education.
- Taylor, S., Fayed, N., & Mandich, A. (2007). CO-OP intervention for young children with developmental coordination disorder. *OTJR: Occupation, Participation and Health*, 27, 124-130.
- Tennant, A., & Pallant, J. F. (2007). DIF matters: A practical approach to test if differential item functioning makes a difference. *Rasch Measurement Transactions*, 20, 1082-1084.
- Three Star Press (2005). School AMPS 2005 computer-scoring software (Version 2005). Fort Collins, CO: Author.
- Townsend, E. A., & Polatajko, H. J. (2007). *Enabling occupation II: Advancing an occupational therapy vision for health, well-being, and justice through occupation*. Ottawa, Ontario: CAOT Publications.
- Tristán, A. (2006). An adjustment for sample size in DIF analysis. *Rasch Measurement Transactions*, 20, 1070-1071.
- UNESCO (1994). *The Salamanca statement and framework for action on special needs education*. Paper presented at the World conference on special needs education: Access to quality. Retrieved 2010-11-12 from http://www.unesco.org/education/pdf/SALAMA_E.PDF.
- United Nation (2002). *World fit for children*. Paper presented at the New York: United nations children's fund. Retrieved 2010-11-12 from <http://www.worldfitforchildren.com/>
- Unsworth, C. (2000). Measuring the outcome of occupational therapy: Tools and resources. *Australian Occupational Therapy Journal*, 47, 147-158.
- Vandenberg, N. L. (2001). The use of a weighted vest to increase on-task behaviour in children with attention difficulties. *American Journal of Occupational Therapy*, 55, 621-628.

References

- Weintraub, N., & Kovoshi, M. (2004). Changing practice patterns of school-based occupational therapists in Israel. *Occupational Therapy International, 11*, 40-51.
- Wilcock, A. A. (1999). Reflections on doing, being, and becoming. *Australian Occupational Therapy Journal, 46*, 1-11.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Vincent, R., Stewart, H., & Harrison, J. (2008). South Australian school teachers' perceptions of occupational therapy reports. *Australian Occupational Therapy Journal, 55*, 163-171.
- World Health Organization [WHO] (2001). *International classification of functioning, disability and health: ICF*. Geneva, Switzerland: Author.
- Wright, B., & Linacre, J. M. (1989). Observations are always ordinal; Measurements, however must be interval. *Archives of Physical Medicine and Rehabilitation, 70*, 857-860.
- Wright, B., & Masters, G. N. (1982). *Rating scale analyses*. Chicago: MESA Press.
- Wright, B., & Stone, M. (1979). *Best test design*. Chicago: MESA Press.
- Wright, B., & Stone, M. (1999). *Measurement essentials* (2nd ed.). Wilmington, DE: Wide Range.
- Wright, B. D. (1995). Which standard error? *Rasch Measurement Transactions, 9*, 436.
- Ziviani, J., & Rodger, S. (2006). Environmental influences on children's participation. In S. Rodger & J. Ziviani (Eds.), *Occupational therapy with children: Understanding children's occupations and enabling participation* (pp. 41-66). Oxford: Blackwell.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores*. Ottawa, Ontario: Directorate of Human Resources Research and Evaluation, Department of National Defense.