

Tiina Lautamo

# Play Assessment for Group Settings

## Validating a Measurement Tool for Assessment of Children's Play Performance in the Day-Care Context

Esitetään Jyväskylän yliopiston kasvatustieteiden tiedekunnan suostumuksella  
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S212  
lokakuun 19. päivänä 2012 kello 12.

Academic dissertation to be publicly discussed, by permission of  
the Faculty of Education of the University of Jyväskylä,  
in Auditorium S212, on October 19, 2012 at 12 o'clock noon.



UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2012

# Play Assessment for Group Settings

Validating a Measurement Tool for  
Assessment of Children's Play Performance  
in the Day-Care Context

JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH 450

Tiina Lautamo

# Play Assessment for Group Settings

Validating a Measurement Tool for  
Assessment of Children's Play Performance  
in the Day-Care Context



UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2012

Editors

Leena Alanen

Department of Education, University of Jyväskylä

Pekka Olsbo, Sini Tuikka

Publishing Unit, University Library of Jyväskylä

URN:ISBN:978-951-39-4890-0  
ISBN 978-951-39-4890-0 (PDF)

ISBN 978-951-39-4889-4 (nid.)  
ISSN 0075-4625

Copyright © 2012, by University of Jyväskylä

Jyväskylä University Printing House, Jyväskylä 2012

To my dear grandchildren, Aava and Oliver  
May your world be full of play and joy!

## ABSTRACT

Lautamo, Tiina

Play Assessment for Group Settings (PAGS): Validating a measurement tool for assessment of children's play performance in the day-care context.

Jyväskylä: University of Jyväskylä, 2012, 69 p.

(Jyväskylä Studies in Education, Psychology and Social Research

ISSN 0075-4625; 450)

ISBN 978-951-39-4889-4 (nid.)

ISBN 978-951-39-4890-0 (PDF)

Yhteenveto

Diss.

The primary aim of this thesis was to evaluate aspects of validity and reliability of the PAGS play performance scale. The PAGS is an observation-based evaluation of the quality of children's play performance when children are playing in groups.

The data were collected by observing children aged 1 year 8 months to 8 years 9 months in free play situations in Finnish day-care centers. In studies I and II the two-faceted Rasch model was used to examine the internal scale validity and the person response processes of the PAGS. In study II differential item functioning (DIF) analysis was also conducted to identify possible group-specific items between children with specific language impairment (SLI) as a potential comparison group for typically developing children (TD). The effect of the differentially functioning items on the mean play performance measurements between the two subgroups of children was controlled with independent samples t-tests. In studies III and IV the three-faceted Rasch model was used to study the rater consistency and severity estimates. In study IV the stability of the individual measures of the children's play performance was also investigated with intra-class correlation (ICC).

The results supported the internal scale validity for the PAGS. We found that the PAGS separated the children with low play performance ability from those with high play performance ability despite the seven differentially functioning items. The majority of the raters scored the PAGS in reliable way. Only slight differences were observed in rater severity. The ICC for single measures of the children's play performance separated by a one-week interval was fairly weak. This indicates that play is a phenomenon that varies from one time to another and cannot be assessed by short-term observation without controlling for environmental effects.

We concluded that the PAGS is a valid tool for identifying children who experience challenges in their play performance, and hence can be used for practical purposes. Overall the results indicated that the raters scored the PAGS in a reliable way, and that they behaved like independent experts. However, the slight variability reveals that DIF should be controlled for when using the PAGS to compare play performance across different subgroups of children. The children's response processes also warrant further study.

**Keywords:** Play Assessment for Group Settings (PAGS), play performance, specific language impairment (SLI), internal scale validity, person response validity, differential item functioning (DIF), inter-rater reliability, test-retest reliability, instrument development

**Author** Tiina Lautamo, LicHSc  
Department of Rehabilitation and Social Sciences JAMK  
University of Applied Sciences  
and  
Department of Education  
University of Jyväskylä  
Vapaudenkatu 85 a 21, 40100 Jyväskylä  
tiina.lautamo@jamk.fi

**Supervisors** Marja-Leena Laakso, PhD  
Department of Education  
University of Jyväskylä

Tuija Aro, PhD  
Department of Psychology  
University of Jyväskylä

Timo Ahonen, PhD  
Department of Psychology  
University of Jyväskylä

**Reviewers** Marja Cantell, PhD  
Department of Psychology  
University of Groningen

Lou Ann Grisworld, PhD, OTR  
Department of Occupational Therapy  
University of New Hampshire, USA

**Opponents** Marja Cantell, PhD  
Department of Psychology  
University of Groningen, Netherlands

## ACKNOWLEDGEMENTS

The years I have spent on this research process have been both challenging and enjoyable. I have had the opportunity work with a subject, the play performance of a child, which is important for me as an occupational therapist, mother and, not least, grandmother. While many other things have happened in my life, both supportive and challenging, this journey has been worth it. This research process, which has enabled me to meet many children personally in daycare environments, has greatly enriched my life. I would especially like to thank, first, all the children, parents, daycare professionals, and therapists who contributed their time during the data collection.

My warmest gratitude goes to my three supervisors: *Marja-Leena Laakso*, my main supervisor, thank you for your whole-hearted engagement, supervision, and support throughout this process. I am grateful that, despite our different theoretical backgrounds, you have respected me as an occupational scientist. We have had fruitful conversations over the similarities and differences in our thinking; *Tuija Aro* and *Timo Ahonen*, the other members of my supervision group, and also co-authors, I thank you both for your help and comments during the different stages of the process. You have been enormously constructive and supportive.

As an occupational scientist it has been valuable for me to be able to reflect on my thinking in the company of my colleagues, other occupational science researchers, and in this way receive valuable feedback. Thank you, *Anna-Liisa Salminen*, *Greta Häggblom-Kronlöf*, *Toini Harra*, *Kristiina Juntunen*, *Maarit Karhula*, *Mari Kantanen*, and *Katja Kanelisto* for always being there.

I wish profoundly to thank all my co-authors, *Anna-Liisa Salminen*, *Anders Kottorp*, *Marja-Leena Laakso*, *Tuija Aro*, *Timo Ahonen*, *Ulf Törmäkangas*, and *Maija Heikkilä* for the countless hours you have spent in helping me in the data analysis and writing processes of our research articles. I gratefully thank *Michael Freeman*, *Jyri Manninen* and *Tarja Karjalainen* for reviewing language of my writings.

Then, my dearest friends, *Paula*, *Tarja* and *Pipsa* and my children, *Janne*, *Toni*, *Jere*, *Minttu*, and *Sofia*, I thank you for supporting me when I has hesitated whether to continue. Thank you for believing in me and encouraging me I needed you. I am grateful to have you around me.

My research work during this journey has been supported by grants from the *Eemil Aaltonen Foundation*, *Finnish Konkordia Foundation*, *Nordisk Forskeutdanningsakademin (NorFa)*, the *Finnish Association of Occupational Therapists*, *The Finnish Foundation for Municipal Development*, and *JAMK University of Applied Sciences*.

I thank all the people who have supported me in their various ways.

Jyväskylä 15.9.2012  
Tiina Lautamo



## FIGURES

FIGURE 1	Play as an occupation: The theoretical frame of a child's socially shared play performance (Lautamo, 2012).....	21
FIGURE 2	The theoretical elements of play in the Play Assessment for Group Settings (PAGS) (Lautamo et al., 2005, revised 2012).....	24
FIGURE 3	Targeting of the play abilities of children: Children's play measures in relation to the item difficulty calibrations of the two groups of children: typically developing and SLI. (differentially functioning items bolded).....	42

## TABLES

TABLE 1	Assessment tools for use in real-life situations.....	19
TABLE 2	The items of the Play Assessment for Group Settings (PAGS)(version 3) and the play elements they outline.....	26
TABLE 3	Comparison of traditional validity concepts and Rasch validity concepts as they apply to the PAGS. ....	29
TABLE 4	Distribution of participants in studies I - IV.....	35
TABLE 5	Validity and reliability evidence collected in studies I-IV .....	36
TABLE 6	The Pags versions used in each study and removed items.....	40
TABLE 7	Items that were relatively more and relatively less challenging for children with SLI .....	49

## LIST OF PUBLICATIONS

- I Lautamo, T., Kottorp, A., & Salminen, A-L., (2005). Play assessment for group settings: A pilot study to construct an assessment tool. *Scandinavian Journal of Occupational Therapy*, 12: 136-144.
- II Lautamo, T., Laakso, M-L., Aro, T., Ahonen, T., & Törmäkangas T., (2011). Validity of the Play Assessment for Group Settings: An evaluation of differential item functioning between children with specific language impairment and typically developing peers. *Australian Occupational Therapy Journal*, 58: 222-230: doi: 10.1111/j.1440-1630.2011.00941.x.
- III Lautamo, T., & Heikkilä, M., (2011). Inter-rater reliability of the Play Assessment for Group Settings. *Scandinavian Journal of Occupational Therapy*, 18:1: 3-12: doi: 10.3109/11038120903480048.
- IV Lautamo, T., Laakso, M-L., Aro, T., Ahonen, T., (submitted *manuscript*). Test-retest reliability of the Play Assessment for Group Settings (PAGS).

The original articles have been reprinted with the kind permission of the publishers.

## CONTENTS

ABSTRACT

ACKNOWLEDGEMENTS

FIGURES AND TABLES

LIST OF PUBLICATIONS

CONTENTS

1	INTRODUCTION .....	13
1.1	Background and rationale .....	13
1.2	Play as an occupation .....	14
1.3	Occupation.....	14
1.3.1	Occupational performance.....	15
1.3.2	Subjective and objective perspectives on play performance ....	15
1.3.3	Occupational competence and identity .....	16
1.4	Top-Down approach .....	16
1.5	Tools to evaluate play performance.....	17
1.6	Development of the Play Assessment for Group Settings (PAGS) ...	21
1.6.1	Theoretical frame of play performance.....	21
1.6.2	Operationalization of the PAGS items .....	24
1.7	The Rasch measurement model – validity and reliability aspects ....	27
1.7.1	The Rasch measurement model .....	27
1.7.2	Aspects of validity.....	28
1.7.3	Aspects of reliability .....	31
2	AIMS OF THIS THESIS.....	32
2.1	The overall aim.....	32
2.2	The specific research aims .....	32
3	METHODS .....	34
3.1	Participants and administration .....	34
3.2	Procedures and data analyses.....	35
4	RESULTS .....	39
4.1	The test content of the PAGS .....	39
4.2	Evidence of validity based on internal structure of the PAGS.....	41
4.3	Evidence of validity based on relations to other variables .....	43
4.4	Evidence of validity and reliability based on response processes....	43
4.4.1	Observed response patterns .....	43
4.4.2	Rater response processes; consistency and severity.....	44
4.4.3	Test-retest of play performance observation.....	45
5	GENERAL DISCUSSION.....	46
5.1	Main considerations related to the test content and internal structure of the PAGS.....	46

5.2	Power to separate children in play performance ability .....	50
5.3	Variability in observed response patterns .....	51
5.4	Rater consistency and severity.....	51
5.5	Reproducibility of the PAGES.....	54
5.6	Other methodological considerations.....	55
	5.6.1 Sample size considerations .....	55
	5.6.2 Limitations of the studies.....	55
	5.6.3 Day care as the evaluation environment .....	56
	5.6.4 Ethical considerations.....	56
6	PRACTICAL IMPLICATIONS.....	59
7	YHTEENVETO .....	61
	REFERENCES.....	63

# 1 INTRODUCTION

## 1.1 Background and rationale

My overall aim in this research project was to conceptualize play performance as an observable occupation in order to validate a tool for measuring this. The outcome, the Play Assessment for Group Settings (PAGS), is intended for observation-based evaluation of the play performance of children in social settings (e.g. day-care) (Lautamo, Kottorp, & Salminen, 2005). My long-term goal was to contribute to knowledge on the play performance of children with mild disabilities, as such children often experience difficulties in their peer relations and play performance in day-care settings. The aim was to construct a valid and reliable tool for use as a true top-down approach to identify children in need of support or therapeutic interventions for optimal participation. The aim of this thesis was to evaluate the validity and reliability of the PAGS measures and scale in order to gather evidence related to the usefulness of PAGS in practice.

The frame of reference in this study is based on the disciplines of occupational science and occupational therapy. Therefore, in the following sections, I first define play as an occupation and in relation to children's occupational participation (1.2), and describe the top-down approach used when the focus is on client-centered practice (1.3). I then give a brief introduction to the four most common and currently available play assessment tools planned for use in real-life situations (1.4). The construction of a new measurement tool always begins with a general idea about what phenomenon one wishes to measure; this in turn demands careful conceptualization of that phenomenon. I therefore introduce my thinking on play performance as a theoretical concept (1.5), as this is at the heart of the idea of the PAGS instrument. Because the research questions focused on instrument development, I briefly discuss the Rasch measurement model and define the terms *validity*

and *reliability* in relation to this thesis, and how they are evaluated from the perspective of the Rasch measurement model (1.6).

## 1.2 Play as an occupation

Children's play has been researched by many disciplines over several decades (e.g., Bundy, 1993; Eisert & Lamorey, 1996; Hartley, 1963; Kooij, 1989; Lieberman, 1977, Shutton-Smith, 1997). Play is seen as an essential part of children's development and way of experiencing the world (Piaget, 1962; Vygotsky, 1978). Occupational therapists have been concerned about children's play as an occupational role, that is, about how children assume their roles as playmates (Burke, 1993). On the other hand, play has also been viewed as an indicator of other abilities, or as a child's developmental level (Eisert & Lamorey, 1996; Knox, 1997; Stagnitti et al., 2000; Kelly-Vance et al., 2002). Occupational science, however, shifts the focus away from a functional view and role theory and explores play as an occupation, describing its essential features and how it is supported by intrapersonal and environmental factors. In addition, it has been argued by Lawlor (2003) that we should shift the focus of our interest away from children's personal challenges toward the study of children as socially occupied beings. Furthermore, if we want to understand play as a social occupation, we need to study play as social performance, not just as "doing", but as a matter of "doing with" (see Lawlor, 2003, p. 426).

## 1.3 Occupation

In occupational science and occupational therapy, the term *occupation* is a core concept. In the Oxford English Dictionary (1989), occupation is defined as "the action of seizing, taking possession of, or occupying space or time." It is also defined as a "series of actions in which one is engaged." Occupation means engagement in the performance of an activity that has meaning and purpose for the person (American Occupational Therapy Association (AOTA), 2008). It is also assumed that engagement in occupation provides structure to everyday life and contributes to health and well-being (Kielhofner, 2002; Wilcock, 1999). Occupation can be seen as participation in different daily life arenas (Law, Steinwender & Leclair, 1998) and can be regarded as a basic human need. Occupational participation is seen as the broadest level of occupation, as it relates to engagement or involvement in a life situation (World Health Organization, WHO, 2001). The term *occupational participation* is used here to refer to engagement in doing that is part of the child's sociocultural context and that is preferred. Play can be understood as a primary and voluntary occupation in which a child participates daily by doing something that matters

with someone else (Lawlor, 2003). Therefore, I use here the term *play* in the meaning of occupation when describing play as an arena of participation.

### **1.3.1 Occupational performance**

The term *occupational performance* pertains to carrying out actions, doing something, and being engaged in doing, for example, playing with one's playmates. Here, I use the term *play performance* to describe a child's observable process of doing, in this case playing. Play performance is shaped by the culture and personal habits of the child. The environment also impacts a child's play performance. Fisher (2009) argues that in order to enact occupational performance, the person must perform a series of observable goal-directed actions over time. When studying play performance, we observe this process of doing rather than an end-product (Rubin, Fein, & Vandenberg, 1983). That is, play tasks and goals are seldom determined beforehand, but children create ever changing play narratives in situ. Play performance becomes observable when a child engages in actions or a series of actions which gradually shape a play narrative. For example, picking up a doll and hugging the doll, putting it to bed, covering it and singing a song to it. In the beginning of the process there may be a socially shared idea of the play, but the story is built up, adapted, and negotiated during the play narrative with the child's co-players. The environment needs to be adjusted to the purposes of the play, and consequently object transformations will be made if necessary during the process. The child uses different learned skills to be able to join this ever changing social process of playing.

### **1.3.2 Subjective and objective perspectives on play performance**

Engagement in doing involves not only objective observable performance but also the subjective experience of the child (see Kielhofner, 2002). The objective perspective can be observed, for example, in whether a child is able to access a play setting, or whether he or she takes an active part in play, or is included in the social activities in that setting (Coster & Khetani, 2008). The subjective perspective on participation can be considered as a sense of belonging (WHO, 2001) or experience of mastery, pleasure and satisfaction (Kielhofner, 2002). The subjective perspective is related to the meaning and importance of taking part in play performance. However, we need to remember that what is observed during play might not directly reflect the child's actual experience of involvement, such as taking part, being included, or engaged in an area of life, or being accepted (WHO, 2001; Ueda & Okawa, 2003). Thus, a child may look uninvolved to an observer, but may experience him- or herself as very involved in a play situation. When play is self-initiated, self-chosen, and self-controlled, we can assume play performance to be meaningful (subjective experienced involvement) to a child (Bundy, 1997; Hasselkus, 2002). However, we need to realize that with objective observation tools we always lose some features of children's play experiences.

### 1.3.3 Occupational competence and identity

Play performance is seen as an important indicator of the occupational competence of the child. Occupational competence depends on experiences of doing things oneself, a sense of causation, emerging interests and values (Christiansen, 1999; Kielhofner, 2002). Children gain their occupational competence - a sense of mastery - in childhood, and gradually developing occupational identity - a sense of who they are - by performing and participating (Kielhofner, 2002). In addition, when we see play as an occupation, we can assume that participation in play is essential to children's feeling of competence, occupational identity and well-being.

Research has shown that children who are challenged in their development also experience challenges in occupational participation. They are approached significantly less often by play partners, participate less in group social play, and engage in more individual play and onlooking behavior (e.g. Liiva and Cleave 2005). Since engagement in occupations is shown to promote health, well-being, and personal identity (Meyer, 1977; Wilcock, 1993; Christiansen, 1999), difficulties in playing or taking part in socially shared situations may give rise to further problems in acquiring the experience, competence and skills needed in daily life (Pellegrini, Dupuis, & Smith, 2007). Therefore, assessing children's play performance should be a fundamental responsibility of professionals working in pediatric rehabilitation and/or in early education. The PAGES offers us a perspective on a child's possible occupational challenges. These challenges need to be understood when planning interventions to enable participation and optimal learning possibilities (Casby, 2003). Consequently, participation in play performance reflects the extent of engagement in the range of activities that accomplish the larger goals of daily life.

## 1.4 Top-Down approach

The past two decades has seen a shift towards a client-centered approach to practice where clients are empowered to have control over their wellbeing and to be engaged in their rehabilitation (Sumsion & Law, 2006). In the case of children, client-centered practice may also be referred to as child-centered or family-centered practice. This view has implications for assessment: therapists need to actively seek the child's point of view as a part of the evaluation process (Sumsion & Law, 2006). The top-down approach facilitates occupation-focused and client-centered practice through evaluating occupational performance in real-life situations and seeking clients' perspectives (Weinstock-Zlotnick & Hinojosa, 2004). When, we focus on the top-down approach in the course of working with children, we need to gain a broad picture of who the child is, and what the child's needs and desires are. We need to know what the child wants to perform and is able to perform. We want to ensure that the child is able to



perform in a manner that brings satisfaction and supports the child's desired level of participation in all the arenas of daily life (see Fisher, 2009). In like manner, it is important to determine challenges in the child's play performance that limit the child's participation. Subsequently we need to clarify and interpret possible causes (personal factors, environment, and/or body functions); however, our main focus is to plan occupation-based interventions that support the child's participation and enable play performance. The focus of evaluations and interventions should be on play as occupation (see Fisher, 2009).

In contrast to the top-down approach, a bottom-up approach has commonly been used which begins with a focus on the evaluation of the child's underlying body functions (see WHO, 2001), specifically those that are thought to cause the child's problems with the performance of daily occupations. The bottom-up approach to assessment emerged from the medical model of practice. This approach contains the danger that we are no longer evaluating the quality of performance but instead focusing the possible impairment of underlying body functions and their remediation, which may not be meaningful to the child and are often isolated from relevant daily life contexts (Brown & Chi-Wen Chien, 2010). This approach might also be time consuming, while there is no guarantee that learned specific skills or remediated body functions will transfer to occupational performance in daily environments (see Fisher 2009).

## 1.5 Tools to evaluate play performance

When I started my research project, instruments which can be used to measure a child's body functions in natural play situations were abundant, but proper tools for the assessment of play performance in real-life situations as such, were lacking (see Bundy, 2001). To my knowledge no validated play assessment tools were—and are still not—in use in the Finnish day-care system, even though, play is considered a primary activity of children and one of the core values in the national curriculum guidelines on early childhood education and care is to secure healthy and safe environments that allow play performance for a child (STAKES, 2004). Furthermore, only a few valid play assessment tools are in use in child rehabilitation in Finland, and none of these has been officially translated into Finnish and culturally validated. Since the processes of translation and culturally validating an existing tool can be time consuming and costly, owing to possible royalties, it is important to find a tool with proven validity that matches the purposes of the assessment. In addition, the chosen instrument should be culturally relevant and measure all the aspects of the phenomenon that are potentially relevant in the practice.

Here, I introduce four of the existing tools planned for use in real-life play situations. I present two of the occupational therapy based tools: the Test of Playfulness (ToP) and the Revised Knox Preschool Play Scale (RKPPS); and, for comparison, two transdisciplinary tools: the Transdisciplinary Play Based Assessment (TPBA-2) and the Play in Early Childhood Evaluation System

(PIECES) (Table 1). Only one of the above, the ToP, focuses on play per se. The three others more or less also focus on the child's capacities and specific skills when describing the developmental stages of play behavior.

Bundy (1997, 2001) developed the Test of Playfulness (ToP) to measure children's play using Rasch analyses in order to construct a linear continuum of more or less playful behaviors based on observations. The purpose is to measure a child's attitude towards play, this is the manner in which that the child approaches play (Bundy, 1993). The ToP reflects four elements of playfulness (originally defined by Lieberman, 1977): intrinsic motivation, suspension of reality, internal locus of control and framing (giving and reading the cues in play) (originally discussed by Baetson, 1971). The measurement of playfulness enables occupational therapists to focus on the complexity of performance within meaningful play activity (Bundy 1997, 2001). Bundy et al. (2001) presented strong evidence of the validity and reliability of the ToP test items. The main limit of the tool is that it is only for the use of occupational therapists. To use the ToP in a reliable way, the user must also have proper training and be a calibrated rater. A computer program to adjust the raw scores of a child's playfulness to the total logit score is not yet available.

While play has been commonly seen as an arena of children's skill acquisition and growth, most of the existing play assessments have focused on players' capacities to play. The Revised Knox Preschool Play Scale (RKPPS) (Knox, 1997) examines play as an overall outcome according to a developmental scheme. The earlier version of Preschool Play Scale (PPS) (Knox, 1974) has been reported to be widely used in occupational therapy practice (Couch, Deitz, & Kanny, 1998). However, several changes have been made in the revised and renamed version. The current version of the KRPPS consist 4 dimensions: space management, material management, pretense/symbolic, and participation, which are divided into 12 categories of play behaviors: gross motor, interests, manipulation, construction, purpose, attention, imitation, dramatization, type, cooperation, humor, and language. The RKPPS provides numerous scores, including overall play age, 4 dimension ages, and 12 category ages. Child play age and play profile provide useful information for planning interventions. The current version of the RKPPS has been reported to be clinically useful in the assessment of children in cases where it is not possible to test them using other developmental standardized tests (Bundy, 2001).

The Transdisciplinary Play Based Assessment-2 (TPBA-2) (Linder, 2008) is the revised version of the TPBA (Linder, 1993) and is intended for the assessment of children from 6 months to 6 years. It is the most widely used and internationally recognized (Kelly-Vance & Ryalls, 2007) instrument. The aim is to assess the child's developmental skills as well as her/his underlying developmental process, interaction patterns, and learning style (Linder, 1993, 2008). The TPBA-2 is meant to be used by a team, comprising professionals and parents. The TPBA-2 provides developmental guidelines to analyze the child's developmental level, learning style, interaction style, adaptive behaviors, and other relevant developmental behaviors. A developmental age score and a functionality score can be established for each of the developmental domains – cognitive, communication, motor, and social emotional.

TABLE 1 Assessment tools for use in real-life situations

	Test of Playfulness (ToP)	Revised Knox Preschool Scale (RKPPS)	Transdisciplinary Play-Based Assessment (TPBA)	The Play in Early valuation System (PIECES)
To be used by	Occupational therapists	Occupational therapists	Transdisciplinary	Transdisciplinary
Play factors	How player approaches play	Player's capacity to play	Player's capacity to play	Player's cognitive capacity to play
Purpose	Captures four elements of playfulness: intrinsic motivation, internal control, freedom of reality, and framing	Provides a developmental description of a child's underlying capacities Limited information over child's play interests	Utilizes a natural play environment for the purposes of assessing underlying capacities, learning style, interaction patterns and other behaviors	Focus is on exploratory and pretend play sub domains
Age range	3 months to 15 years	0 to 6	0 to 6	0 to 7
Setting	Natural settings, both outdoors and indoors	Natural settings, both outdoors and indoors	Natural play environment	Natural play environment
Methods	Observation in at least two different 15- to 20-minute sessions	Observation in two 30-minute sessions	Observation in several structured and unstructured sessions. Total time 60 to 90 - minutes	Observation in play situations, 30 to 40 – minutes
Scoring	Criterion referenced	Criterion referenced	Criterion referenced	Criterion referenced based on normative data
Interpretation of results	Not yet available commercially, no standard scores available. Until then, qualitative interpretation	Total play score is calculated by averaging dimension scores	A developmental age score and a functionality score	Developmental criteria based on other studies / typical development, discrepancies in developmental age
Evidence on reliability and validity	Item response validity Person response validity Rater reliability (Bundy et al., 2001) Clinical utility (Cameron et al., 2001)	Validity and reliability studies only performed with the earlier PPS version (Beldsoe & Shepherd, 1982; Harrison & Kielhofner, 1986). Preliminary inter-rater reliability and criterion validity (Jankovich et al., 2008)	Social validity (Myer et al., 1996) Criterion validity (Kelly-Vance et al., 1999)	Inter-rater reliability Test-retest reliability (Kelly-Vance & Ryals, 2005) Criterion validity Clinical utility (several studies done but not published) (Kelly-Vance & Ryals, 2005)

The Play in Early Childhood Evaluation System (PIECES) (Kelly-Vance & Ryalls, 2005) is the most recently developed transdisciplinary approach. The PIECES also grew out of empirical work originally based on Linder's TPBA cognitive development assessment guidelines. It is intended for children under the age of six or seven. The PIECES assessment involves observation of a child engaged in free play, and the assessment can be conducted in any setting with any set of toys as long as the toy set is large and varied enough to elicit a wide range of behaviors. The PIECES coding guidelines differ from those of the TPBA as they include a 13-item sequence that makes up the exploratory/pretend play core subdomains. The PIECES focuses wholly on the cognitive development of a child. It is an observation-based tool used during free, independent play and it can be completed in any childhood setting. The information obtained is compared to the norms of typically developing children to determine areas in need of intervention. The resulting PIECES guidelines have been evaluated empirically and found to have high inter-rater reliability (Kelly-Vance & Ryalls, 2005). In addition, this high inter-rater reliability was achieved with a relatively simple training procedure. That is, individuals with a background in observational techniques can be trained to accurately and reliably use these guidelines to assess play behavior with as little as half a day of training (Kelly-Vance et al., 2007).

The four assessment tools described above have their advantages and disadvantages. The RKPPS revised version has not been validated. Validity and reliability studies have mainly been performed with the earlier PPS version (Beldsoe & Shepherd, 1982; Harrison & Kielhofner, 1986). Jankovich et al. (2008) recently compared children's play age to their chronological age and inter-rater agreement between two raters. They recommended changes to the guidelines and also further examination of the psychometric properties of the RKPPS. The TPBA-2 has been argued to be expensive in terms of time demands and therefore limited in applicability in early education and rehabilitation contexts (Stagnitti, 2009). When I started developing the PAGES, the ToP was under development and the PIECES had not yet been published. One reason of developing the PAGES was that I had the ambitious goal of combining the elements of playful attitude and developmental stages of play in order to assess their manifestation in the child's play performance while still making the instrument relatively easy to use by all professionals working with children.

## 1.6 Development of the Play Assessment for Group Settings (PAGS)

In response to the need for an instrument that all professionals working with children would find easy to use and that (at the same time) would provide objective knowledge on a child's play performance, I constructed the Play Assessment for Group Settings (PAGS), originally in Finnish. Since a good theory is essential for developing any assessment tool, and the construction of a measurement tool always begins with an idea about what is to be measured, the first phase in the development of the tool was to conceptualize play theoretically as observable performance.

### 1.6.1 Theoretical frame of play performance

Here, the theoretical frame of reference of *play performance* is conceptualized through three main elements affecting the dynamic process of play: (1) the child's *spirit*, and (2) the child's *skills*, and (3) the child's *environment* (Figure 1). To describe in more detail what needs to be observed in the relations between the elements of the child's spirit and skills, the elements of the environment have been conceptualized further. These elements are named as (4) *meaningful doing*, (5) *mindful doing*, and (6) *expression of mastery* (Figure 2).

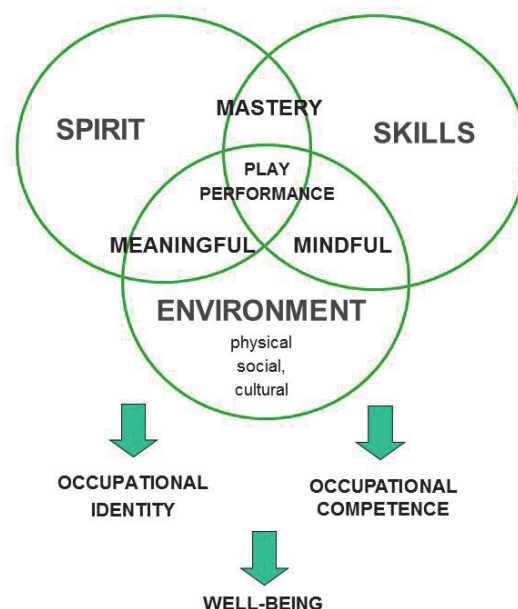


FIGURE 1 Play as an occupation: The theoretical frame of a child's socially shared play performance (Lautamo, 2012)

*Spirit* is seen here as the truest self of the child, something she or he is attempting to express through actions (Egan & DeLaat, 1994). The term 'spirit' reflects the concepts of inner drive and self-actualization which can be observed as the excitement, confidence and effort that the child brings into a play situation when she or he finds just the right level of challenge in the process of doing (Ayres, 1972; Bundy & Murray, 2002; Christiansen, 1999).

*Skills* are defined here as learned actions a child is performing while participating in daily activities. Within play performance we can observe a number of discrete mindful and goal-directed actions. These observable actions that a child performs refer to skills within doing (Fisher, 2006; Forsyth & Kielhofner, 1999). According to Sturges (2009), an important question is whether the play performance in question demands specific skills. It is evident that a child needs to have both developmentally derived skills and skills that make an event playful. Play needs also skills to imagine and skills to socially share play reality, that is the ability to read the minds of other individuals (Farrant et al., 2006). This means not only the ability to use symbols as basic images (Casby, 2007), but also the ability to evaluate actions on the basis of one's own or others' intentions, beliefs or plans (Nelson et al., 2003). According to Harris (1992; 1996) both of these socio-cognitive abilities demand a process of setting aside one's current point of view and imaginatively sharing another's perspective. Language skills profoundly facilitate the development of these kinds of skills because conversation and complex use of language with repairs and clarifications involve a constant exchange of differing points of view. Play is not only valuable and pleasurable for its own sake but it is also an important arena for optimal social skill acquisition. In addition, in a supportive social environment the child dares to take risks and might perform on a higher skill level than in solitary play. Play performance on this proximal level of development is also fundamental in the process of skill acquisition (Vygotsky, 1976), not only communication and interaction but also motor and process skills.

The *environment* challenges a child to participate in his or her life arenas. The environment offers physical and social opportunities, resources, constraints and demands. Whether a child notices these or whether they influence a child's behavior are dependent on his or her skills, experiences, values, interests, and habits (Kielhofner, 2002). If the environmental challenges are too demanding for the child, she or he will be unable to cope in the situation and may withdraw from the play or just remain an onlooker. Therefore, how environmental factors affect the play situation should always be taken into consideration. Moreover, it should be remembered that the environment is not only physical or social but it is also interpreted and shaped by culture (Altman & Chemers, 1980). In other words, we need to understand how, in particular, cultural and social factors affect the way children play.

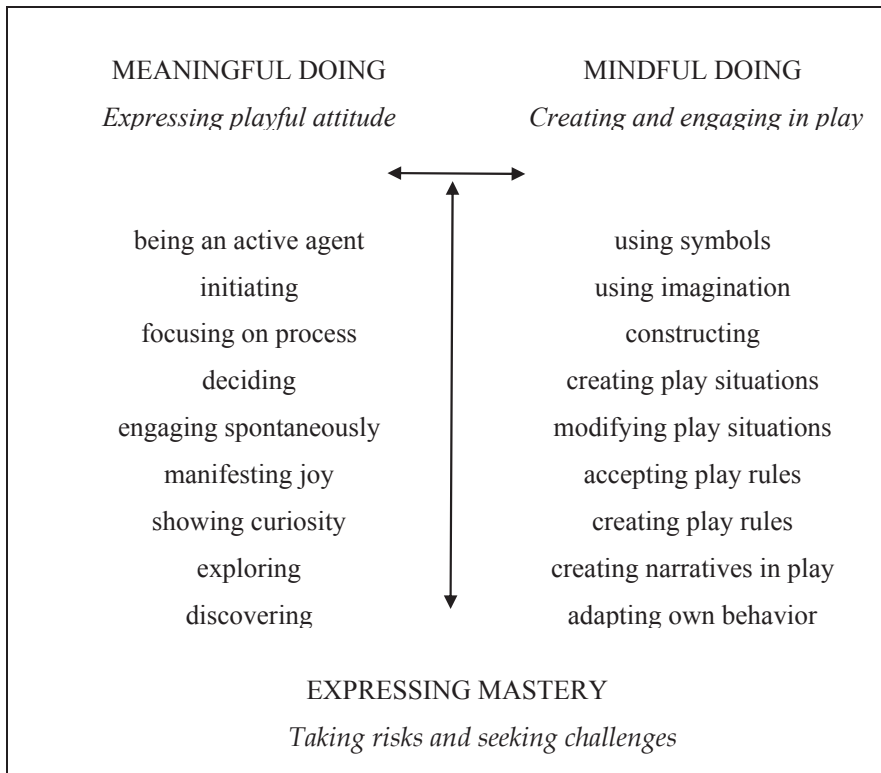
*Meaning* pertains to the significance of the play performance to the child. Engagement comes from a sense of meaning derived from or experienced during doing (Fisher, 2009). When the environment presents an appropriate level of challenge to the child's skills and interests, then the child experiences

*meaningful doing* and expresses a playful attitude in play. By playing, children gradually learn how they can cause things to happen, with the result that the desire to have an effect on the environment becomes a strong motive and manifestation in their play (Bundy, 1997; Ferland, 1997; Kielhofner, 2002). Understanding the elements of a playful attitude enables us to focus on the complex phenomenon of meaningful play performance (Lieberman, 1977; Neumann, 1971; Bundy et al., 2001). When children express this playful attitude, they, for example, express emotions, show curiosity, explore, are spontaneously engaged, and exhibit a sense of humor during play (Bundy, 1997, Bundy et al., 2001; Ferland, 1997; Lieberman, 1977). In the other words, meaningful doing becomes observable when the child approaches the social and physical play environment with playful actions which reflect a playful attitude.

*Mindful doing* focuses on the process of doing where the child uses her or his skills in a way which enables more than just automatic functioning (Langer, 1989). It is the process of doing where children welcome and create new information and new categories in their play world, and become aware of others' points of view (Hasselkus, 2002). They gradually gain experiences with the minds of self and others through social performance (Symons, 2004). By playing, children acquire the ability to pretend and understand the pretending of others (Leslie, 1987). The aspects of mindfulness focus on how a child can adapt her/his own behavior and adjust to the environment in order to share the play narrative with other players.

*Mastery* is attained through the experience of doing things for one self, a sense of causation, emerging interests and values (Christiansen, 1999; Kielhofner, 2002). Children want to master challenges, and their interests reflect their expanding skills. In order to participate and to become a competent player in all play arenas, a child needs not only gradually improving skills, but also to maintain her/his occupational identity (Kielhofner, 2002). Feeling and expressing mastery in turn is dependent on the occupational identity of the child and further enhances the child's actual experience of involvement and participation.

FIGURE 2 The theoretical elements of play in the Play Assessment for Group Settings (PAGS) (Lautamo et al., 2005, revised 2012)



### 1.6.2 Operationalization of the PAGS items

The next step was careful operationalization of the theoretical elements of play performance to the observable play performance items. The first version of the PAGS contained 54 items; however, after careful theoretical and statistical analysis we ended up with 38 items (version 3) (Table 2). The items in the PAGS are meant to represent the latent trait of the child's play ability and were constructed on the basis of the above-described dynamically interwoven theoretical aspects of play performance: meaningful doing, mindful doing, and expressing mastery.

Meaningful doing can be observed mainly with items 1 to 13, 15, and 16; mindful doing by items 17 to 38; and expressing mastery by items 12 and 14 (see table 2). Meaningful doing was operationalized, for example, to assess play behavior where a child expresses curiosity towards the environment and objects (item 2) or decides on his or her own actions independently rather than imitates others (item 8). Mindful doing was operationalized, for example, to assess play behavior where a child comes up with his or her own ideas for play (item 20) or



actively modifies or adapts the theme as the play progresses (item 34). There are also items clearly describing both meaningful and mindful doing, such as when a child takes part in shared play as an active participant (item 9) or adapts his or her actions in order to make the play more challenging or enjoyable (item 13). An example of expressing mastery is when a child expresses the feeling of competence by, for instance, describing accomplishments or expressing contentment (item 12).

In order to measure children with less ability and those with more ability in play, the items were designed to represent a developmental continuum from less to more demanding play performance (see Bond & Fox, 2007; Wright & Stone, 1979). The developmental continuum of play performance was designed to measure 2- to 7-year-old children's play performance. To cover this age range, there are easy items, not-so-easy items, more difficult items and even more difficult items. For example, the items 1 (enjoys activity and play; he or she has fun playing) and 3 (begins a play without the support of an adult) are the easiest and items 14 (takes risks and seeks challenges) and 16 (playfully breaks the rules of the activity and looks for challenges in order to accommodate the play) are the hardest items for all children. This means that we expect the child to be more likely to obtain higher scores on the easier than harder play items, and that the easier play items will be more likely to be easier for all children than the harder play items (Wright & Stone, 1979). This item hierarchy is expected to be relatively stable irrespective of the group of children being evaluated.

The PAGES items are scored according to the frequency of specific play behaviors a child enacts in a social context. Scores are recorded on a 4-point scale indicating the relative amount of time that a child's "doing" reflects that item (1=hardly ever; 2=seldom; 3=often, and 4=nearly always). On the score sheet, percentage approximations of the values indicated by these Likert scores are given to help the scorer, but are only offered as rough guides and not meant to be applied literally. The use of time-related scoring is based on fact that as children move up from one phase of play performance to the next, they continue to use old play patterns along with the new ones (Reilly, 1974). It is important to note that the Likert scale scores are never precise measures. They are estimations about how much the child's play performance reflects a certain item. Therefore, for to research the validity of the PAGES scorings, we needed to conduct proper statistical analyses, such as the Rasch measurement model (see Bond & Fox, 2007).

TABLE 2 The items of the Play Assessment for Group Settings (PAGS)(version 3) and the play elements they outline.

The items of the PAGS
1. enjoys activity and play; he or she has fun playing (manifesting joy)
2. expresses curiosity towards the environment and objects (showing curiosity, exploring)
3. begins a play without the support of an adult (initiating)
4. devotes him or herself to the process of playing and enjoys the activity (engaging spontaneously, focusing on process)
5. explores possibilities offered by the environment (exploring, discovering)
6. takes part in new plays with curiosity (being an active agent, deciding)
7. plays tricks and makes jokes (has a sense of humour)
8. decides on his or her own actions independently rather than imitates others (deciding, being an active agent)
9. takes part in shared play as an active participant (sharing reality, being an active agent)
10. embraces activity in new surroundings (exploring)
11. expresses feelings during the play (expressing emotions)
12. expresses the feeling of competence by, for instance, describing accomplishments or expressing contentment (expressing mastery)
13. adapts actions in order to make the play more challenging or enjoyable (manifesting joy, adapting own behaviour)
14. takes risks and seeks challenges (risk taking, expressing mastery)
15. teases in a positive and playful manner (manifesting joy, has a sense of humor)
16. playfully breaks the rules of the activity and looks for challenges in order to adapt the play (modifying situations, creating play rules, risk taking)
17. plays in organized manner and his or her play has obvious goal (being an active agent, creating play situations, adapting own behavior)
18. accepts other child's play acts as part of shared play (sharing reality, accepting play rules)
19. attributes qualities to objects, e.g. the doll is ill, the car is broken (using symbols)
20. comes up with his or her own ideas for plays (using imagination)
21. uses diverse and varying toys and objects while playing (exploring, discovering)
22. replaces missing objects with imaginary ones, e.g. pretends there is food in the pot (using symbols)
23. shares toys unity with other players (accepting play rules)
24. learns/understands new play rules (accepting play rules, adapting own behavior)
25. builds play surroundings like houses or huts (constructing)
26. acts a theme or a story (using imagination, creating narratives)
27. invents plays about everyday situations like being at home or in a shop (creating play situation)
28. names a role for him or herself, plays at being someone else (using imagination)
29. uses objects symbolically, e.g. a box as a table, a block as a car (using symbols)
30. uses objects in a creative and original way (using symbols, discovering)
31. contributes to the ongoing play theme adapting its rules (shares reality, creating narratives, modifying situations)
32. invents adventure plays like being a princess, a pirate or a spaceman (using imagination, creating narratives)
33. describes what is going on while playing (sharing reality)
34. actively modifies or adapts the theme as the play progresses (modifying play situations, creating narratives)
35. discusses the rules of a play with other playmates (creating play rules, negotiating ideas)
36. moves from one play event to another easily in collaborative play (adapting own behaviour, sharing reality)
37. invents new plays and shares his or her ideas with playmates (using imagination, sharing reality)
38. understands the play rules others have set (sharing reality)

## 1.7 The Rasch measurement model – validity and reliability aspects

### 1.7.1 The Rasch measurement model

In observing play performance, ordinal scores are used to rate performance in accordance with specific criteria. The structural limitations of ordinal data are widely known. Ordinal data are expressions of observed qualities and cannot be treated as measurements (Wright & Linacre, 1989, Bond & Fox, 2007). Since almost all the traditional statistical techniques commonly used (e.g. mean, SE, correlation coefficient) are based on interval data, the underlying assumptions of those statistics are violated when applied to ordinal data. Such applications may also yield results that will induce misleading interpretations (Bond & Fox 2007). In response to the awareness of the limitations of traditional psychometric statistics when applied to ordinal data, Rasch measurement models have become widely used in constructing and validating tests in rehabilitation and education (e.g., Fisher, 1994, 2006; Bundy et al., 2001; Bond & Fox, 2007).

When we observe play, we see a manifestation of a latent trait (Andrich, 1988). Both factor analysis and Rasch analysis could be used to test whether a data set is indicative of a single latent trait (Sick, 2011). However, factor analysis only identifies closeness to the underlying variable, but not the item's location on the scale. Rasch analysis takes as its starting point the assumption that a set of items is intended to measure a single construct (Bond & Fox, 2007). More importantly, Rasch analysis provides both item and person location on the scale. Rasch analysis assumes that items intended to measure a latent trait should be initially developed on a conceptualized scale from easy to hard, and that in the present case children's play performance should reflect the items from this perspective. Rasch analysis of the data tests this concept, and a linear scale can be created based on a mathematical measurement model with consistent units (logits), in this case for the play performance abilities of the children, the easiness or hardness of the play items, and rater severity.

The assertions of the Rasch measurement models are probabilistic in nature. This means, that individuals' performance on an item can never be predicted with 100% accuracy. However, the Rasch perspective requires that the data representing a latent trait should fit the MFR model. When using MFR methods in the development of an instrument, items that do not fit the basic assertions of the model must be considered for omission from the instrument or reworded. The MFR model provides the necessary objectivity for the construction of a scale that is separable from the distribution of the attribute across the persons it measures and provides approximations of measurements that help us understand the processes underlying the reason why items, persons or raters behave in a particular way (Bond & Fox, 2007).

In order to evaluate the validity of the PAGS scale and the reliability of both the children's play performance ability measures and rater response processes, using a modern test method, a Rasch measurement model was deemed appropriate (Bond & Fox, 2007). The Rasch model used in this thesis is the many-faceted Rasch (MFR) measurement model.

### 1.7.2 Aspects of validity

In developing an instrument or a measurement tool, the most fundamental consideration is validity. Validity refers to the degree to which the empirical evidence and theory support the interpretations of test scores. Commonly, validity is seen as a unitary concept and includes (1) *validity evidence based on test content*, (2) *validity evidence based on internal structure*, (3) *validity evidence based on relations to other variables*, and (4) *validity evidence based on response processes* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). It is important to note that the definitions of validity and reliability in the MFR model differ from those traditionally used (Bond & Fox, 2007; Stemler, 2004). A brief summary of the Rasch validity types employed in this thesis in relation to the current validity concept is presented in Table 3.

Validity is not a property of the instrument, but of the instrument's scores and their interpretations (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). An instrument's scores will reflect the underlying construct more or less accurately but never perfectly (Cook & Beckman 2006). The process of validation involves collecting evidence to provide a sound scientific basis for the proposed interpretations of the scores.

(1) Validity evidence based on test content involves evaluating the "relationship between a test's content and the construct it is intended to measure." (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). The theoretical test content should represent the whole construct. Consequently, we look at the definition of the construct, the intended purpose of the instrument, the process for developing and selecting items, and the wording of individual items (Haynes et al., 1995). Content evidence is presented here through a series of steps taken to ensure that the items represent the play performance construct.

(2) Scores intended to measure a single construct should yield a homogenous internal structure (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). MFR analyses provide item fit statistics, that is, indicators of how well each item fits within the underlying construct (Bond & Fox, 2007). When estimating a scored phenomenon, in this case play performance, with the MFR model, all the test items are expected to form an unidimensional continuum representing the theoretical construct of the phenomenon being measured (Bond & Fox, 2007; Tesio, 2003). The existence of this unidimensional

continuum of the test items can be confirmed by reasonable fit statistics. Unidimensionality is important, since

lack of measurement uniformity may lead to incorrect estimates of effects in research, and hence also in decision making at the individual level (Decker et al., 2005). Items are arranged in ascending order of difficulty (Bond and Fox, 2007).

TABLE 3 Comparison of traditional validity concepts and Rasch validity concepts as they apply to the PAGS.

<i>Traditional source of validity evidence</i>	<i>Rasch validity types</i>
<i>Validity evidence based on test content</i>	<i>Similar</i>
<i>Validity evidence based on internal structure</i>	<i>Internal scale validity</i> <i>Item goodness of fit</i> <i>Differential item functioning /stability of item hierarchies across the groups</i>
<i>Validity evidence based on relation to other variables</i>	<i>Validity evidence based on relation between generated person ability measures and other variables</i> <i>Separation index (separation reliability)</i>
<i>Validity evidence based on response processes</i>	<i>Person response validity (observed response patters of the person)</i> <i>Targeting test item difficulties to ability sample</i> <i>SE mean</i> <i>Rater response processes</i> <i>Rater severity</i> <i>Overall percentage of unexpected ratings</i>

Another aspect of validity based on internal structure is to ensure that the instrument is free of differential item functioning (DIF). DIF analysis provides an indication of unexpected behavior by an item in a measurement instrument. DIF occurs when people from different groups have a different probability of getting a particular score on a test item (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). Systematic variation in responses to specific

items (DIF) among subgroups that were expected to perform similarly suggests a flaw in the internal structure of the instrument, whereas confirmation of predicted differences provides supporting evidence for this aspect of validity. When using MFR measurement methods, we expect the items to retain their order of difficulty regardless of the person being evaluated.

(3) Validity evidence based on relations to other variables can be confirmed by a good correlation between the present scores and those obtained with another instrument or outcome for which such a correlation would be expected (Cook & Beckman 2006), for example, the correlation between play performance scores and developmental delays (e.g. specific language impairments). An important aspect of validity in relation of other variables is that the instrument is sensitive enough to be used to identify differences between groups that are expected to differ on the trait being measured. This kind of evidence supports the practical and clinical utility of the proposed test for determining if a child needs special support or a therapeutic intervention to facilitate play performance. In MFR analysis, this reflects the separation ability (separation index) of the test.

In addition, the Rasch measurement model can be used to convert (a person's) (persons') raw ordinal scores for a set of items into ability measurement expressed in an equal-interval, log odds probability unit termed a logit. These logit measures are the interval units and can be used in a more consistent way in further statistical analyses, for example, for comparison of the differences between groups (Bond & Fox, 2007).

(4) Validity evidence based on response processes represent the behaviors required to respond to an item. In constructed-response items, the individual's response process is the primary focus and is intended to reflect a range of performance dimensions. In MFR analyses, each person's ability measure, each item's difficulty, and each rater's severity calibration can be positioned on a common unidimensional scale of measurement. MFR analyses provide person scores and rater scorings fit statistics, that is, indicators of how well each person's scores or each rater's scorings fit and are targeted within the underlying construct of the test content (Bond & Fox, 2007).

One benefit when using MFR analyses is that we can easily count the overall percentage of unexpected scorings and also interpret the possible causes if any systematic bias emerges among the scorings. Furthermore, observation-based assessments always require the judging of raters, since an understanding of a person's response processes also contributes to the validity evidence collected for the test score interpretations. When raters rate a performance or complete a checklist, the raters' precise responsibilities should be outlined in the rating specifications. The influence on scores of judges and raters weakens measurement reliability, which in turn impacts validity. When assessment administration procedures are more standardized, there is less opportunity for raters to influence the person's scores (Cook & Beckman, 2006).

### 1.7.3 Aspects of reliability

Reliability is a necessary, but not sufficient, component of validity (Cook & Beckman, 2006). Reliability refers to the stability or consistency of a measurement when a test procedure is repeated on an individual or on the same group of individuals (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). Some variations are always present between raters and repeated measures. An instrument that does not yield reliable scores does not permit valid interpretations, and real changes cannot be distinguished. Nevertheless, reliabilities are often reported as though they were invariable characteristics of tests. However, they depend not only on the construction of the test, but also on the distribution of the examinee sample tested (Fisher, 1992).

Two forms of reliability that are typically considered in assessment concern rater reliability. Rater reliability refers to the consistency and severity of scores that are assigned by two independent raters (inter-rater) and that are assigned by the same rater (test-retest) at different points in time. To investigate test-retest reliability we calibrated the children's play performance responses, that is, to determine whether the measures remained stable within the chosen time interval. We tested inter-rater reliability by using MFR measurement methods and test-retest reliability in a more traditional way (ICC, one-way ANOVA).

Estimation of inter-rater reliability through the use of the MFR model (Linacre, 1994) allows rater severity to be determined using the same scale as person ability and item difficulty. Moreover, the difficulty of each item and the severity of all the raters who rated the items can also be directly compared. The MFR approach also allows one to evaluate the extent to which each of the individual raters is using the scoring scale in a manner that is internally consistent. In other words, even if raters differ in their own definition of how they use the scale, the fit statistics will indicate the extent to which a given rater is faithful to his or her own definition of the scale categories across items and people (Stemler, 2004). We expected the raters to be within a reasonable severity range. Specifically, rater severity is independently estimated and transformed into a linear measure of the underlying factor. Rater severity measures are useful for estimating the extent to which systematic differences in severity exist between raters. In addition, when using the MFR model, differences in rater severity can be controlled for at both the individual and item level in the algorithm that computes the participant ability estimates (Bond & Fox, 2007).

## **2 AIMS OF THIS THESIS**

### **2.1 The overall aim**

The general aim was to conceptualize play performance as an observable occupation in order, first, to construct and, second, to evaluate the validity and reliability of the measures and scales used in the Play Assessment for Group Settings (PAGS) and hence its practical usefulness as an instrument for evaluating children's play performance in day-care settings.

### **2.2 The specific research aims**

*Validity evidence based on test content*

- To evaluate validity evidence based on the test content of the PAGS (Study I and II)

*Validity evidence based on internal structure*

- To evaluate validity evidence based on the internal structure of the PAGS by estimating if the items of the PAGS work together to define a single construct that can be used to measure children's play performance in group settings (Study I and III)
- To evaluate validity evidence based on the internal structure of the PAGS by determining if the PAGS is free of differential item functioning (DIF) when used to evaluate the play performance of two groups of children: children with specific language impairment (SLI) and their typical developing (TD) peers (Study II)



*Validity evidence based on relations to other variables*

- To evaluate validity evidence based on relation to other variables by determining if the PAGES has the power to differentiate children who have been observed or diagnosed to have challenges in their play performance (Study I and II)

*Validity evidence based on response processes*

- To evaluate validity evidence based on children's response processes (Study I)
- To evaluate rater consistency and severity, when scoring children's play performance (Study III and IV)
- To evaluate the test-retest reliability of the PAGES by calibrating the children's play performance responses with a one-week interval between repeat scorings (Study IV)

## 3 METHODS

### 3.1 Participants and administration

The participants of this thesis were children varying in age from 1.8 to 8.9 years. They were either (1) typically developing children (with no observed challenges in their development), or (2) children with observed or diagnosed challenges (mainly with specific language impairment (SLI), but also other diagnoses, e.g. ADHD, Asberger, CP) and (3) raters who were graduates or students in the early education or rehabilitation disciplines. An overview of the participant demographics in the four different studies is presented in Table 4. During the four studies, we collected the data on the play performance scorings of 208 Finnish children; this database was used to model the MFR for the PAGES.

The children's play performance data for Studies I and II were collected in 33 different day-care centers in Central Finland, both in cities and rural areas, but in fairly similar cultural environments. Early education professionals scored familiar children's play performance after reading brief descriptions of the PAGES theory and item scoring instructions. The participants (12 raters) in study III were professionals (early education and rehabilitation) enrolled on a continuing education program on play therapy interventions. These raters co-scored 6 video cases and 6 self-selected live cases from among their own clients. The participants (6 raters) in study IV were bachelor degree-level occupational therapy students. They co-scored (in pairs) 29 typically developing children twice at a one-week interval in two different day-care centers.

TABLE 4 Distribution of participants in studies I - IV

Study	I	II	III	IV
<b>Number of participants</b>				
Children	93	110	78	29
Raters	Several	several	12	6
<b>Age range (yrs)</b>	2.1 to 8.9	3.0 to 6.5	1.8 to 8.8	3.0 to 5.5
<b>M (SD)</b>	5.4 (1.3)	5.4 (0.87)	5.3	4.5
<b>Gender</b>				
Boys	50	32	38	12
Girls	43	23	40	17
<b>Diagnostic groups</b>				
Typically developing	70	55	30	29
Observed or diagnosed challenges	23	55	42	
<b>Data overlap in studies</b>	Partly in study II	Partly in study I	None	None

### 3.2 Procedures and data analyses

In the studies that comprise this thesis, we have mainly used the MRF model when collecting evidence on whether the play ratings of the children contributed in a meaningful way to the underlying latent trait of play performance. In studies I and II we collected validity evidence related to the test content, internal structure and relations to other variables. In studies I, III and IV we collected validity evidence based on response processes (Table 5).

Initially, in all four studies, we implemented MFR analyses using the FACETS computer program (Linacre, 1987-2007) to generate a linearized play performance measurement for each child. The FACETS program summarizes all unexpected responses (residuals) in the observations into goodness-of-fit statistics and also generates the play performance ability value (logit) for each child. These analyses calibrated the quality scores of play performance of each child, the difficulty of each item, and the severity of each rater. In order to be able to compare the item functioning of two subgroups in study II, DIF analysis was performed to generate group-specific play item difficulty calibrations. This specific MFR analysis was completed by using WINSTEPS ([www.winsteps.com](http://www.winsteps.com)), version 3.63.0 (Linacre, 2005).

In studies I and II we used two-faceted MFR analyses (children and items) and in studies III and IV we used three-faceted MFR analysis (children, items, and raters). In addition, the data analysis in studies III and IV were driven by anchored item values in order to generate statistics on how the reduced dataset fitted the expectations of the existing MFR model for PAGES. Anchoring refers to a procedure that specifies that present values are to be used in the analysis (Linacre, 2011). The anchoring values were obtained from the MFR analysis of

the total database of the play performance scorings of 208 Finnish children. We also used the Statistical Package for the Social Sciences (SPSS) computer program to perform descriptive and inferential statistical analyses in all four studies.

TABLE 5 Validity and reliability evidence collected in studies I-IV

Validity and reliability evidence	Study I	Study II	Study III	Study IV
Validity evidence based on test content	Theoretical considerations within data analysis	Theoretical considerations within data analysis		
Validity evidence based on internal structure	Evaluation of the uniformity of the child's play performance scale (MFR, <i>MnSq</i> , <i>SEm</i> )	Evaluation of stability of skill item hierarchies (MFR, DIF)		
Validity evidence based on relation to other variables	Evaluation of the power to differentiate play performance between children (MFR, separation index)	Evaluation of differences between groups expected to differ (independent samples <i>t</i> -test)		
Validity and reliability evidence based on response processes	Evaluations of children's response processes (MFR, <i>MnSq</i> , <i>SE mean</i> )		Evaluation of rater response processes (MFR, consistency and severity) <i>SE mean</i>	Evaluation of rater response processes (MFR, consistency and severity) <i>SE mean</i>
			Overall percentage of unexpected ratings	Evaluation of test-retest responses of the individual children (ICC ANOVA)
				Overall percentage of unexpected ratings

When investigating *the validity evidence based on the test content*, theoretical consideration was given to the wording of the PAGES items in relation to the statistical analyses and internal structure of the PAGES. *The validity based on internal structure*, in turn, was studied by generating mean square (*MnSq*) and standardized (*z*) goodness-of-fit statistics. We used both measures of fit: the infit and the outfit. These statistics were used to evaluate acceptable response patterns for the play items (studies I and II) that fit the assertions of the Rasch model for the PAGES. The criteria for an acceptable goodness of fit (both infit and outfit) of the play items were  $MnSq \leq 1.5$  (study I) and  $\leq 1.3$  (study II) and  $z < 2$  (Wright & Linacre, 1994). The infit is the weighted mean-squared residual which is sensitive to unexpected responses near the point where decisions are being made, while the outfit is the unweighted mean-squared residual and is sensitive to extreme scores. It is generally expected that 95% of the items should meet the criterion that 5% may misfit by chance at  $z < 2$ .

In order to study the stability of the skill item hierarchies, differential item functioning (DIF) analysis was conducted (study II) to estimate the relative difficulty of each item separately for the TD children and children with SLI. In order to answer the question whether there were play items that were relatively easier or more difficult for the children with SLI than the typically developing children, the item calibrations were plotted using a 95 % confidence interval (CI). Furthermore the DIF probability (MantelHenzel) for meaningful differences between the two groups was set at  $< 0.05$ . Our basic assumption was that the play item difficulty calibration hierarchy would remain stable across different groups. Generally, it is expected that 95% of the items would meet this criterion, that is, just one or two items would lie outside the CI (Bond & Fox, 2007).

In order to evaluate if the PAGES had the power to separate the children who had observed or diagnosed challenges in their play performance we studied the separation index and separation reliability for items and persons. A separation index  $< 2.0$  and a reliability coefficient  $> 0.80$  were expected (Fisher, 1992)

In order to investigate *the validity evidence based on relations to other variables*, we investigated whether the two groups of children (TD and SLI) differed in their play performance measures. To investigate if differentially functioning items affect the ability of the PAGES to separate children who have challenges in their play performance, two different independent samples t-tests were conducted (study II): first, using the play performance values (logits) of the individual children with all 38 play items (the PAGES, Current version 3) (Lautamo, 2009), and second, using the performance values (logits) of the individual children with the limited set (28) of uniform and stable PAGES items. The effect sizes for the two independent groups were also calculated (Cohen's *d*) (Rosnow & Rosenthal, 2003). Equalities of variance were examined using Levene's test. Age and gender effects were controlled for by matching the sub groups. In study II, the effect sizes for the two independent groups were

calculated based on standardized mean differences (Cohen's  $d$ ) (Rosnow & Rosenthal, 2003).

To evaluate *the validity evidence based on the response processes* of the PAGES we investigated person response processes (study I) and rater response processes (studies III and IV). These response processes were studied by generating mean square ( $MnSq$ ) and standardized ( $z$ ) goodness-of-fit statistics for the children and the raters. The criteria for an acceptable goodness of fit (both infit and outfit) of the children's play performance and the raters were set at  $MnSq \leq 1.5$  and  $z < 2$  (Wright & Linacre, 1994). It is generally expected that 95% of the items should meet the criterion that 5% may misfit by chance at  $z < 2$ . Mean standard error estimates ( $SE$ ) for the separation of the children (study I) and raters (study III and IV) were also investigated. The size of the  $SE$  is influenced by how well the data fit the model assertions and how well targeted the difficulty of the items and severity of the raters is to the abilities of the individuals (Bond & Fox, 2007). The criterion for an acceptable standard was  $SE \leq 0.30$  (Tham, Bernspång, & Fisher, 1999).

In order to investigate inter-rater reliability (rater response processes) and test-retest reliability (studies III and IV) the first step was to determine whether the raters were consistent enough, that is, whether their scorings of the children's play ability fitted the MFR model for the PAGES. In order to evaluate inter-rater reliability we also used another method adopted from a previous study (Coto, Fisher, & Mayberry, 1995). We counted the overall proportion of unexpected scorings of raters. Based on the rate of unexpected scorings for the total database ( $n=208$ ), we set our criterion for an acceptable level of individual unexpected scorings at  $\leq 4\%$ , when  $z = 2$ . To examine whether the raters behaved like independent experts we calculated the Rasch Kappa value, which indicates the relation between the expected and the observed percentage agreement between the scorings of the raters. In general, we expect the observed percentage agreement to be close to the expected percentage agreement. On the whole, the value of the Rasch Kappa index is expected to be positive and close to 0.0 (Linacre, 2011).

We expected the raters to maintain the same relative *severity*. According to Linacre (1994), calibration stability within  $\pm 0.30$  logit is the best that can be expected for any variable. However, for practical purposes, Linacre (1994) states that variation up to  $\pm 0.50$  is likely to be free from bias. Thus, we expected variation between raters to be within  $\pm 0.50$  logits. Furthermore, to confirm consistency, the separation index of the raters was evaluated and expected to be  $< 2.0$  (Wright 1996).

To study the test-retest reliability (study IV) of the PAGES, we converted ordinal raw scores based on five consistent raters' scorings of children's play behaviour to play performance measures (logits) for two different observation occasions scored at a one-week interval. To examine the test-retest reliability of the PAGES individual measures scored at a one-week interval, we performed intraclass correlation coefficient (ICC ANOVA) analyses. To confirm the test-retest reliability, we expected  $p < 0.05$  and the ICC value to be, for excellent reliability,  $> 0.75$ , and, for moderate reliability, within the range 0.40 to 0.74 (Fleiss 1986).

## **4 RESULTS**

### **4.1 The test content of the PAGES**

When investigating the other aspects of validity, theoretical consideration was given to the content of the PAGES in relation to the statistical analyses. Originally The PAGES contained 54 theoretically constructed play items. After critical consideration, 51 of the items were included in the first analysis in study I. In the step-by-step analysis in study I, four of the items that did not demonstrate acceptable patterns were removed. The final analysis then revealed that the 47 play items of the PAGES (version 1) defined a single unidimensional construct. After Study I, the development work on the PAGES was continued by leaving out two items which described the same play behavior as some other remaining item and at the same time were on the same level of the challenge hierarchy. In Study III, we used the revised version (version 2) of the scale with 45 items. In study II, we used same procedures as in study I to examine the fit of the items and reduced the number of items to 38 (version 3). Studies II and IV were conducted with this third version of the PAGES. In Table 6 you can find the information which version was used in each study and the removed items. The third version of the PAGES is introduced earlier in chapter 1.5.2.

TABLE 6 The Pags versions used in each study and removed items

Study	I	II	III	IV
<b>Scored PAGES items and version</b>	54 items Version 1	38 items Version 3	45 items Version 2	38 items Version 3
<b>Items used in final analyses</b>	51 to 47 items	35 in DIF anal. 38 and 28 items in log transformation for t-tests	38 items	38 items
<b>Misfit items</b>	<b>Retained items :</b>  Child brakes the rules in way that brings more fun into play	<b>Retained misfit items:</b> Teases in a positive and playful manner  Accepts other child's play acts as part of shared play  <b>Item with few scorings:</b> Understands the play rules others have set	None	Not applicable
	<b>Removed 7 items in step by step analysis:</b>  3 theoretically un relevant items removed before data analysis  Plays with conventional toys, e.g. dolls and toy cars  Participates skillfully in running, skipping and climbing  Appropriately follows others' play without disturbing and destroying  Appropriately chooses solitary or social play	<b>Retained 7 DIF items:</b>  Explores possibilities of the surroundings  Embraces activity in new surroundings  Learns/understands new play rules  describes what is going on while playing  Discusses the rules of play with other playmates 36 moves from other event of play to another easily in collaborative play  Invents new play and shares his/her ideas with playmates	<b>Removed 7 items before analysis:</b>  Shows interest towards new play  Participates play enjoying the act  Participates spontaneously  Accepts help from the peer in play  Takes own role  Acts familiar theme e.g. from fairytales or films  Brakes the rules in way that brings more fun into play	Not applicable



## 4.2 Evidence of validity based on internal structure of the PAGES

Evaluation of the internal scale validity of the PAGES has been continuous process and confirmed in studies I, II and III. In Study I revealed that the 47 play items defined a single unidimensional construct such that the 46 (97.9 %) items demonstrated acceptable response patterns ( $MnSq \leq 1.5$ ,  $z < 2$ ) of the underlying construct of the PAGES. All the scale items were well distributed along a linear scale and targeted the play performance of children. In the early phases in the study III we used same procedure that was used in study I and we found 38 well-functioning items. In study II we used 38 items version of the PAGES. While we used more strict criteria ( $MnSq \leq 1.3$ ,  $z < 2$ ) for acceptable response patterns of items we found two items to misfit. The two misfit items; (a) teases in a positive and playful matter, and (b) accepts other child's play acts as part of shared play, seemed to be sensitive to environmental effects (common behavior rules in day-care). Moreover, removing these items did not influence the fit statistics of the other items. Therefore we decided to retain them in the construct of the PAGES (version 3).

Anyhow, when analyzing the differential item functioning (DIF) in study II we used only the 35 items that were well-functioning and had a required number of scorings. We found 28 (80%) of the 35 items to be stable across both groups of children (SLI and TD). However, seven (20%) items differentiated between the SLI and TD groups in terms of the challenge they presented to these children (Table 6). This was more than we expected. Although the item hierarchy in the PAGES was largely stable, group-specific play items emerged (Figure 3). Two items were relatively less challenging for the children with SLI than for their TD peers. Five items were relatively more challenging for children with SLI than their TD peers. Four of the five relatively more challenging items clearly required expressive or receptive language skills. We decided to retain for now those seven items in the scale, the reasons are discussed more detailed later (chapter 5.1).

<i>Measures (logits)</i>	<i>Child</i>	<i>PAGS play items, TD</i>	<i>PAGS play items, SLI</i>
	Higher skills	Harder items	Harder items
+ 6	xx		
+5	x		
+4	x xxx		
+3	x x xxxx xx x xx		16
+2	xxx x xxxxx xxx xxxxxx xx	16	<b>35,36,37</b> <b>33</b> <b>24,34</b> 13,14,30 31
+1	xxxx xxxxxxx xxxx xxxxxx xxxxxxx xxxxxx		22,26,28,29,32 27 20,23 12 21,25
0	xxx xxx xxxx x xxx xxx xxxx	29 27,31,32 23,30 25,28,34, <b>36</b> 13,21, <b>33,35,37</b> <b>5,10,19,22</b> 11,12,20,26	9,19 6,7,11 <b>10,17</b> 4 8 3, <b>5</b>
-1	xxxxx x xx x x x	6,9 2,8, <b>24</b> 4 7,17 3	2 1
-2	x xx		
-3	x	1	
	Lower skills	Easier items	Easier items

FIGURE 3 Targeting of the play abilities of children: Children's play measures in relation to the item difficulty calibrations of the two groups of children: typically developing and SLI. (differentially functioning items bolded)

### 4.3 Evidence of validity based on relations to other variables

When evaluating the PAGES ability to separate children who have challenges in their play performance we found in Study I the separation index of the children's ability measures to be 4.95 (reliability = 0.96). This indicated that the 47 items of the PAGES separated the children into at least five different levels of ability in their play performance (Fisher, 1992). In addition, the PAGES differentiated the children who had observed or diagnosed challenge from typically developing ones. 18 of the 23 children with observed or diagnosed challenges were among the lower half of all children on the play performance continuum, while the age mean of the disabled children's group was higher than in the typically developed children's group. This pattern was also present in the larger dataset while the separation index of the children's ability measures (n=208) was 4.51 (reliability = 0.95) (not reported).

The ability to separate two groups of children was confirmed in study II. Despite the seven items functioning differentially across the two groups of children (SLI and TD), we confirmed with two different independence sample t-tests that the play performance ability means of the two age- and gender-matched groups of 3.1- to 6.5-year-old children (SLI, n=55, and TD, n=55) differed significantly. In the first analysis of differences in play performance ability (conducted with 38 items) independent samples t-tests confirmed that the means differed significantly ((108),  $t = 5.80$ ,  $p < 0.01$ ), and that the effect size was large (Cohen's  $d = 1.11$ ). In the second analysis of the differences in mean play performance ability (conducted with 28 stable items), independent samples t-test confirmed that the means still differed significantly ((108),  $t = 5.02$ ,  $p < 0.01$ ), and that the effect size was also large (Cohen's  $d = 0.97$ ). Levene's test confirmed the equality of variances in both analyses. Based on these two analyses, we concluded that despite the differentially functioning seven items, the PAGES (current version 3) measures were sensitive enough to discriminate between groups expected to differ (i.e., children with SLI and TD peers).

### 4.4 Evidence of validity and reliability based on response processes

#### 4.4.1 Observed response patterns

When examining the observed response patterns in the children's play performance scorings in Study I, 84 of the 93 (90.3%) children's play performance scores demonstrated acceptable patterns. The other nine children's scorings thus comprised slightly more than the expected 5%, and hence further analysis was undertaken to evaluate if any systematic bias in observed response patterns were found among these children. Three of the nine children (33.3%) had observable or diagnosed challenges in the development. This was only

slightly higher than the proportion overall. Five (55.6%) of the nine children were boys and four were girls. This was the same as in the misfit observed response patterns, and also overall. Six (66.7%) of the nine children were about six years of age (5.10-6.9 years), while the proportion of the same age group in the total group was 28.3%. This pattern was also found in the larger dataset (n=208, not reported). Acceptable response patterns in play performance scorings were observed in 88.1 % of the children.

In spite of the higher proportion of unexpected observed response patterns in children's play performance, the mean SE values of the children's play performance measures were acceptable (in Study I, SE mean=0.26, and in the larger (n=208) sample, SE mean=0.29). This implies that the data accord with the measurement model and indicates that the test item difficulties were well targeted at the play performance ability of the sample. Therefore we concluded that the observed response patterns of the children's play performance did not violate the validity of the PAGES, but must be studied further.

#### 4.4.2 Rater response processes; consistency and severity

To answer the question addressed in studies III and IV of whether the raters were consistent enough, we examined if they conformed the expectations of the MFR model with respect to goodness of fit. In study I, 11 of the 12 raters (91.7%) demonstrated acceptable goodness of fit. In study IV, after removing one too lenient rater's scorings, the remaining five raters fitted the data ( $MnSq \geq 0.5$  or  $\leq 1.5$ ). The mean SEs of rater severity in studies III and IV were acceptable (SE mean = 0.08, and SE mean= 0.06), which supports the model assertions.

The proportion of individual unexpected scorings in study III was 4.0% ( $z = 2$ ), as expected. However, in study IV, the proportion of individual unexpected scorings was 5.6 % ( $z = 2$ ), which was slightly more than expected (< 4.0 %). In fact, in study IV, three specific items explained 48.3 % of the unexpected responses: item 7; plays tricks and makes jokes (16.9 %); item 19; attributes qualities to objects (11.6 %); and item 22; replaces missing objects (19.8 %). These individual unexpected scorings are discussed later.

When analyzing the expected agreement of the rater scorings in studies III and IV, the rater scorings were relatively close to the expectations of the MFR model (0.0). In study III, the Rasch Kappa value was, 0.12 and in study IV, 0.17. This indicates that the raters in this study were scoring the PAGES items in a reliable way and as independent experts (Bond & Fox, 2007).

To examine whether the raters maintained the same relative severity, we investigated the rater calibration values. In study III, the rater calibration values ranged between -0.24 and + 0.27 logits, where the twelve raters fitted within  $\pm 0.51$  logits, and in study IV, after omitting the one too lenient rater, the rater calibration values ranged between -0.24 and + 0.29, where the five raters fitted within  $\pm 0.53$  logits, while the mean of rater severity was centered at zero. This was slightly higher than the expected value ( $\pm 0.50$ ) (Linacre, 1994). The

separation index was 2.09 in study III and 2.93 in study IV, both of which values were slightly higher than the expected value of < 2.0.

#### **4.4.3 Test-retest of play performance observation**

When we investigated the test-retest reliability of the PAGES at a one-week interval five consistent raters were observing totally 29 typically developing children in pairs. We found the intra-class correlation coefficient (one-way) in single measures to be 0.481, (95 % CI 0.15 - 0.72,  $p < 0.01$ ) and in average measures to be 0.649 (95 % CI 0.26 - 0.83,  $p < 0.01$ ). These ICC values indicate that the test-retest reliability on the group level was good, but on the individual level fairly weak.

## **5 GENERAL DISCUSSION**

While the primary aim of this thesis was to evaluate aspects of the validity and reliability of the PAGES play performance scale and the children's play performance observations, my long-term goal was to generate a play performance assessment tool for practical purposes. When the PAGES is used to evaluate children's play performance in group settings, its practical usefulness is determined according to whether early education or rehabilitation professionals can use the results to identify children who are in need of support in play performance, an individual rehabilitation intervention, or consultation services. Furthermore, the usefulness of the PAGES also requires that the measurement tool can be used to evaluate changes and document the effectiveness of interventions. These issues ultimately pertain to the validity of the PAGES evaluations and scales and whether professionals can make reliable interpretations of the quality of children's play performance. I will therefore now discuss the evidence for the validity and reliability of the PAGES play performance measures.

### **5.1 Main considerations related to the test content and internal structure of the PAGES**

In order to provide validity evidence based on the test content and internal structure of the PAGES, we conducted step-by-step MFR analyses. In the present instance the idea that play performance can be assessed as a single phenomenon needed proper theoretical support (see Bond & Fox, 2007). Overall, good item fit supported the internal scale validity of the PAGES. This means that the items of the PAGES worked together and measured the same latent trait, children's play performance. Our statistical findings, therefore, indicate that we succeeded in reaching the goal of capturing the most potentially observable elements of the latent trait of play performance in the same continuum. In addition, this revealed that all three theoretical features—meaningful doing,

mindful doing, and mastery – are captured in the same theoretical continuum of play performance.

Parallel to the data collection and statistical analyses, we collected additional feedback on the contents and wording of the items from the professionals scoring the PAGES. Where necessary, we clarified the instructions for scoring and made slight changes in the wording of the items. We ended up with the current third (38 items) version by omitting and rewording items in step-by-step considerations of statistical and theoretical processes.

Validating an instrument is a fundamental and continuing process that is required to support meaningful and relevant interpretation of test scores in relation to the theoretical structure being assessed. It should be remembered that validity is a characteristic of the interpretation, not the instrument per se, and that the validity of interpretations is always a matter of extent (Cooch & Beckman, 2006). This is especially true of play performance, which is a complex and varying phenomenon in which the challenges for the child differ according to the social and physical environment (Sutton-Smith, 1997). It could be argued that it is impossible to construct a tool that will measure all the possible aspects of play. We accept, therefore, that we are likely to have missed some elements of play. However, the aim was not to attain a perfect realization of the theoretical elements but to capture the most potentially observable elements of the latent trait of play performance.

However, despite the good item fit, the slightly higher than expected variation in the observed response patterns in the children's play performance was reason to further evaluate the evidence based on the internal structure of the PAGES. We investigated the differences in item functioning (DIF) between two subgroups of children. We decided to assess the play performance scorings of children with specific language impairment (SLI) as a potential comparison group for typically developing children (TD), since the former have been found to show less than optimal development in play performance (Casby, 1997; Rescorla & Goossens, 1992). As expected, we found that the children with SLI differed in total play performance from their TD peers. Furthermore, we found seven differentially functioning items (Table 7). The five items were relatively more challenging and two items were relatively less challenging for the children with SLI. Four of the five relatively more challenging items clearly required expressive or receptive language skills. Of these items, three required the adequate use of expressive language and the fourth one required receptive language skills. In other words, all four items related to the difficulties the children with SLI experienced in inventing new ideas and, in particular, sharing them with others by describing what is going on during play. These considerations are in accordance with previous studies which have focused on children's participation in peer relationships and have emphasized the role of language in play in social settings (Fujiki et al., 1999; Liiva & Cleave, 2005).

The fifth relatively more challenging item, item 36, describes the child's ability to understand and be flexible in the ever-changing social reality of play. Earlier studies showed that children with SLI are delayed in controlling and adjusting their behavior (McCabe & Marshall, 2006; Picone & McCabe, 2005) and have difficulties in 'reading the minds' of other individuals, and 'seeing' things from another person's perspective (Farrant et al., 2006). This kind of play performance requires not only language skills, but also the ability to plan and to adapt to the process of doing. We should also take into account aspects of conversation with peers when assessing children's play performance. Where a child is scored down on these particular items, we can conclude that the child has language-related difficulties in play performance and justify seeking a more precise consultation with, e.g., a speech therapist.

The two less challenging items indicated that children with SLI spend relatively more time on activities related to exploring their environment. Our clinical experience, along with earlier findings, suggests that the dominance of the exploring activity could reflect the child's inability to play at a more symbolic level (see Rescorla & Goossens, 1992). On the other hand, deficits in language skills and inability to fluently join in social play (Fujiki et al., 1999) could encourage the child to spend more play time on activities related to exploring the environment. The detailed analysis of the DIF items supported the idea that assessment of play performance with the PAGES reveals problems in play that are in accordance with the child's specific challenges. Therefore, in future studies we need to investigate whether other play patterns and relatively differently functioning items are present among other subgroups of children, for example children who have more social or motor challenges in play, such as children with ADHD, or children from a different cultural background.



TABLE 7 Items that were relatively more and relatively less challenging for children with SLI

Challenge for children with SLI	Reflects on	DIF items	Authors comments
Relatively more challenging items:	The expressive language-related items	33 describes what is going on while playing 35 discusses the rules of play with other playmates 37 invents new play activities and shares ideas with playmates	The difficulties the children with SLI in inventing new ideas and, in particular, sharing them with others by describing what is going on during play
	The receptive language related item	24 learns / understands new play rules	
	Flexibility in social situations	36 moves easily between stages of play in collaborative play	Capability to understand and to be flexible in the ever-changing social reality of play, delayed in controlling and adjusting their behavior
Relatively less challenging items:	Meaningfully explores the environment	5 explores the opportunities provided by the surroundings 10 embraces activity in new surroundings	Behavior that dominates during infancy but is also an important part of play, especially in new environments and situations and reflects the style the child adopts in approaching the environment

## 5.2 Power to separate children in play performance ability

It is important to have a valid assessment tool that can differentiate persons with limitations in their performance from those without limitations. The PAGES items can be used to identify children who are challenged in their play performance. As the single hierarchical continuum of 38 items of the PAGES proceeds from less challenging items to more challenging items, we were able to confirm the following two assertions of the MFR model: (a) a child is more likely to obtain higher scores on the easier play items than harder play items; and (b) the easier play items are more likely to be easier than the harder play items for all children. Based on our results, we confirmed that the PAGES had the power to separate children who have challenges in their play performance, in the present instance children with observed or diagnosed challenges in their development (study I and II).

The power to separate children into different levels of play performance was calculated with a separation index, which is the number of statistically different performance strata that the test can identify in the sample. The separation index of the PAGES varied from 4.51 to 4.95 in the different MFR analyses. This indicated that the items of the PAGES separated the children into at least five different ability levels in their play performance (see Fisher, 1992). In addition, the results of study I indicated that the PAGES measures were related to the other challenges, observed or diagnosed, in the children, and hence provided further evidence of the separation ability of the PAGES.

The results of study II also indicated that despite the differentially functioning of seven items in the PAGES, it succeeded in differentiating the children with SLI from their TD peers in play performance. Because lack of uniformity in measurement may lead to incorrect estimates of the effects in research (see Decker et al., 2005), we conducted two different analyses to determine whether omitting the differentially functioning items affected the mean difference between the two subgroups, children with SLI and their TD peers. In the first analysis, the calibrations of the measurements of the children's play performance were conducted with all 38 items of the PAGES (Lautamo, 2009), while in the second analysis we used the calibration data on the 28 uniformly functioning items. No notable differences were found between the results of these two analyses. The results indicated that the PAGES has the potential to detect differences in children's play abilities, not only in the case of children who have language-related challenges in their play performance but also among children with other challenges. However, the slight variability detected in the uniformity of the PAGES suggests that the DIF items should be controlled for in future studies when comparing play performance between different subgroups of children.

### 5.3 Variability in observed response patterns

The observed response patterns in children's play performance were empirically tested to determine whether their scores responded in an acceptably predictable way, given the expectations for hierarchical ordering of the MFR model for the PAGES (see Bond & Fox, 2007; Wright & Stone 1979). That is, we expected the children's play performance scores to conform to the following two assertions: (a) the child is more likely to obtain higher scores on the easier play items than harder play items, and (b) the easier play items are more likely to be easier than the harder play items for all children. The children's scores on the PAGES items that conform to these expectations will demonstrate acceptable goodness of fit with the Rasch model for the PAGES. However, examination of the observed response patterns revealed that approximately 10 % of the children's play performance scores did not demonstrate acceptable patterns. We considered the possible causes, and concluded that there might be some systematic differences among some subgroups of children or that some raters did not score the items adequately.

As mentioned earlier, the language-related aspects of play performance, especially among the children with SLI, might be one explanation. Another possible explanation concerned the children aged six years or older. After examining the data in study I in more detail, we found a higher proportion of children six years of age among the misfitting patterns. This may indicate that some of the response patterns for the PAGES in this age group can exhibit unexpected variations. However, the limited number of participants in study I did not support an in-depth subgroup analysis to investigate whether there was any statistically significant pattern among the children who did not demonstrate an acceptable observed response pattern. When we further studied (study II) possible subgroup variation, we found seven differentially functioning items across the two groups of children (SLI and TD). The DIF items and the higher proportion of children under six years of age revealed that there might also be differentially functioning items among other subgroups of children. However, the play performance of children can also vary from one environment to another and from one situation to another. Children's play style (Stagnitti, 2009) can also vary and affect raters' scorings. It can be harder to observe skillful play acts, e.g. when a child actively modifies or adapts the theme, during boys' rough and tumble play than when children are playing more calmly and are negotiating play scenes.

### 5.4 Rater consistency and severity

A possible reason for the higher than expected number of misfitting observed response patterns in children's play performance, could be that some raters did

not score the items adequately. Therefore, in studies III and IV we focused on rater response processes by examining rater consistency and severity.

Our results on rater consistency indicated that the majority of the raters scored the PAGES in a reliable way. Eleven of the twelve raters' scorings in study III and five of the six raters' scorings in study IV showed good fit with the assertion of the MRF model for the PAGES. Further analysis was undertaken to ascertain whether the scorings of the uncertain raters exhibited any systematic patterns. In study III we found that the most of one uncertain rater's unexpected individual scores were among her scorings for the video data, while her scorings for live observations were reliable. We concluded that she scored in a reliable way in live situations. The use of videotapes has its limitations, for example, the information available to the rater may be restricted and some aspects of play performance might be difficult to identify (see Hermansson, et al., 2006). In study IV, while no systematic error pattern for specific raters was found, three specific items explained 48.3 % of the unexpected responses (item 7, 19, and 22). These items were well functioning items in the total database of the MFR model for the PAGES, and therefore we concluded that the present data collection was affected by some special concern. Children were scored down in these three items in comparison to their other play performance scores and total play performance level. Possible reasons for these low scores were that opportunities to show this kind of play behavior were lacking during the observation occasions or that the raters had not received clear enough instructions on how to observe this kind of play behavior. The low scores recorded for item 7 (plays a trick and makes jokes), could reflect the effect of the presence of an unfamiliar observer in the play situation or a fragile overall ambience towards playfulness. In the case of item 19 (attributes qualities to objects, e.g. the doll is ill, the car is broken) and item 22 (replaces missing objects with imaginary ones, e.g. pretends there is food in the pot), which measures symbolic actions of the child in play, it might be that the fairly short observation period did not include object substitution or the attribution of object qualities, despite the fact that child is capable of making symbolic use of objects. On the other hand, the toy repertoire available to the child can also affect the amount of symbolic play; however, this was not controlled for in our study.

Our main focus in study IV was on children's play performance responses at a one-week interval, and hence we wanted rater data that were comparable. Since the two raters had co-scored 10 children's play performance together in both time frames, the leniency of one rater seemed to cause some bias to the co-raters fit statistics. One rater was found to be extra lenient (rater 2), and therefore her scorings were removed from the data. When her data were removed, the earlier slightly misfitting scorings of other rater fitted the data. It seems that even if all the raters score the PAGES in a consistent way, variation in the severity of their ratings can also cause bias in co-ratings. When evaluating the intended or unintended consequences of an assessment, previously

unnoticed sources of invalidity can be revealed (Cook & Beckman, 2006); in this case the invalidity seemed to be due to single rater leniency.

When we evaluated rater severity in our studies (III and IV), we obtained rater calibration values of between 0.51 and 0.53. Linacre (2011) stated that for all practical purposes, variation up to  $\pm 0.50$  logits is likely to be free from bias. He also stressed that as sample size increases, the differences between raters become smaller. We used relatively small sample sizes in both studies. Nevertheless, in both studies the variation in leniency between the raters was only slightly higher than expected. MFR analysis also provides a separation index for raters. Our findings revealed that the raters' scorings separated the raters into only two different severity levels. Calibration of rater severity is not needed for practical purposes; however, we recommend controlling for rater severity when using the PAGS for research purposes. Furthermore, when the MFR model is used, possible larger differences in rater severity can be adjusted for on an individual and item level if needed.

Linacre (1989) believes that the phenomenon of rater variation is an expected component of the scoring process. He asserts that raters cannot be trained to achieve similar levels of severity. For example, Mansoor & Houman (2011) and Weigle (1998) confirmed that training reduced rater unfairness but did not eliminate it in terms of severity. Therefore, rater training does not necessarily force raters into exact agreement with each other, but rather trains raters to be self-consistent. This view of the function of rater training allows for some variability in rater variation, seen as a natural part of the rating process, in scoring. Play is acknowledged to be a complex activity that contributes significantly to all aspects of the development of children. The complexity of play demands special skills from the persons observing and scoring children's play performance in social and physical environments (Ferland, 1997). We can assume that day-care professionals working with children on a daily basis are the best professionals to observe children's play performance. However, this is not self-evident; Sutton-Smith (1997), for example, indicated that female teachers can easily misinterpret the noisy and aggressive play of boys. Both the teacher's attitude towards play and the institutional context of day-care affect how play is manifested in children's groups.

Furthermore, while the proportion of unexpected scorings by individual raters was higher than expected in study IV, we did not find any systematic error on the part of any specific rater. This might be caused by slight uncertainty among the inexperienced raters (students) compared to the professionals in early education and rehabilitation who collected the data in the previous waves (Study I, II, and III). According to Weigle (1998), before training, inexperienced raters tended to be both more severe and less consistent in their ratings than experienced raters. Although our inexperienced raters were educated to use the PAGS, self-inconsistency of the raters might nevertheless be one cause of the weak ICC. The possible slight uncertainty shown by the ratings of the inexperienced raters in study IV compared to the study III ratings by professionals (in early education and rehabilitation), would indicate that rater

training cannot totally compensate for earlier experience in observing play performance. We therefore concluded that raters need practical experience and structured instruction to score and reflect on play performance from an occupation-based perspective in order to administer and score the PAGES items in a valid and reliable manner.

## 5.5 Reproducibility of the PAGES

In study IV, we evaluated the test-retest reliability of the PAGES by calibrating the children's play performance responses at a one-week interval. In contrast to other reliability estimates, the test-retest reliability or reproducibility of a test captures not only the measurement error of an instrument, but also the stability of the construct being measured (Schuck, 2004), in this case children's play performance. Play performance can be seen as changeable rather than stable, owing to the strong effect of environmental issues, both social and physical. According to the raters' observations during the data collection, a child's play performance depended considerably on environmental issues; for example, if a child's best playmate was absent, the child may not have wanted to play at all. Mood also affected the attention paid by the child's towards the environment; for example, a more tired child preferred to play alone and concentrate on construction play, while in the second observation the same child was very eager to join a more complex social play scenario. These observations reflected the fact that the three theoretical elements (spirit, skills, and environment) of the theoretical frame of reference of play performance work together in interaction and cannot be separated during play assessment.

Our results revealed that individual children's play performance measures varied when their test-retest responses were evaluated at a one-week interval. The fairly weak correlation between the two measures can be seen as a threat to the repeated measures reliability of the PAGES. On the other hand, in our study a pair of occupational therapy students evaluated a child's play performance during a single 45-minute observation occasion. One week later, the same child was observed again but only one of the raters was the same. It is possible that because only one of the paired raters observed the same child twice, the resulting slight variance in rater severity might have weakened the child's play performance score over time.

Eight children's play performance measures varied by more than 0.50 logits and six varied by more than 1.0 logits across the one-week time interval. According to Linacre (1994), calibration stability within  $\pm .30$  logits is the best that can be expected for any variable, and for all practical purposes variations up to  $\pm .50$  logits would be likely to be free from bias. However, Linacre (1994) also stressed that calibration estimation error is dependent on the size of the sample used to calibrate the items, persons, or raters. For pragmatic reasons associated with time constraints, we used a relatively small sample: 29 calibration cases scored on 35 items. For sample sizes of this magnitude, Linacre

(1994) proposed the criterion of 95% confidence of calibration stability within  $\pm 1.0$  logits. While Linacre (1994) argued that 30 cases for 30 items might be enough for stable measures, he recommended an empirical dataset in the center of a distribution of 100 to obtain more reliable findings. In accordance with these arguments, we can consider our small sample to be stable within  $\pm 1.0$  logits. This indicates that a poor ICC will not necessarily be the case in larger samples. Therefore, there is a need to reevaluate the test-retest reliability with a larger sample, controlling for rater severity, to be able to investigate further whether the PAGES can be used as a reliable tool to measure changes in children's play performance.

## **5.6 Other methodological considerations**

### **5.6.1 Sample size considerations**

The sample size in all our studies was relatively small, and in studies I and II the samples were drawn partly from same database. In study I, in the pilot phase of the scale validity study, we had only 93 children. However, the MFR model for PAGES was updated with a larger database of 208 children. According to Linacre (1994), a sample size of 64 to 144 is needed to ensure 95% and a sample size of 108-243 to ensure 99 % confidence for stable skill item difficulty estimates based on a criterion of  $\pm 0.50$  logits.

In order to equate the age- and gender-matched groups of children in study II, we selected all the available data from our previous database and increased the group size to 55 children per group, although for DIF analyses the recommended sample size varies from 100 (Tristan, 2006) to 200 participants per group (Tennant & Pallant, 2007). Tristan (2006) discusses the fact that DIF analyses depend on the standard error estimates of the items, which are reduced in larger samples.

### **5.6.2 Limitations of the studies**

The main limitation of studies I and II was that only the internal aspect of validity was controlled for. External aspects, such as rater severity or inter-rater reliability may also affect the results of these studies. Since we used several raters to code the data, each of whom made only one or two observations, it was not possible to control for rater severity. The raters had different levels of training in studies I and II for scoring the PAGES, which might have caused a risk for impaired inter-rater reliability. Nevertheless, in study III we indicated that the PAGES scorings between the diverse professional raters were free from bias and comparable with each other. On the other hand, in study IV, our raters were second-year occupational therapy students, whose scorings, despite their rater training, showed slight uncertainty compared to the earlier ratings by professionals (in early education and rehabilitation). In studies I, II, and III the

raters were aware of the diagnoses of the children; this knowledge might have affected the scores of these children (Learn, 1996).

The raters were allowed freely to choose the duration of their observation time in studies I, II, and III, in order to obtain enough information to measure the child's play performance. This might have affected their judgments. Brentnall, Bundy and Kay (2008), suggest that a longer observation time is not necessarily more representative than a shorter time. In their study, in which children's playfulness was scored with the Test of Playfulness, the mean scores of different raters based on 30-minute videotaped observations were significantly lower than those of either the first or last 15 minutes of the same videotaped observations of other raters. However, when investigating if lengthening the observation time might yield scores for items earlier scored "not applicable", they found that in 30-minute observations the only item that was scored in more cases than previously was pretending (Brentnall et al., 2008). On the other hand, the PAGS contains more items measuring pretending in play.

### **5.6.3 Day care as the evaluation environment**

Environmental features (physical, social and cultural) affect the play performance of children. Our purpose was to guarantee the familiarity of the children with the environment by choosing to study children who had been in the same day-care group for a longer period. Other environmental aspects were not controlled for. It has been argued that the play ability of a child should always be observed and measured in natural settings (Bundy, 2001) which are safe and familiar enough to support play performance and enable a playful attitude (Pellegrini et al., 2007). Environments that challenge a child's capacities tend to evoke involvement, attentiveness, and maximal activity (Csikszentmihalyi, 1990), although it must be recognized that natural peer settings are not necessarily equally challenging environments for all children and that the level of challenge can change from day to day. In our experience play performance is also dependent on how the social environment—in the present case day-care professionals—value play as such. In addition, the data in this study were collected among a culturally homogeneous population (no children from other ethnic cultures); this can also have an impact on how the features of play performance are viewed. Play has many universal features, but cultural concepts can also affect interpretation and scoring.

### **5.6.4 Ethical considerations**

Before embarking on the data collection of the studies included in this thesis, permissions (to implement the doctoral thesis proposal) were obtained from the ethical committees of the Central Hospital of Jyväskylä and the University Hospital of Kuopio. Prior to the commencement of Study I, II and IV, ethical approval was also obtained from the managers of the Social and Health Services in the cities where the children attended day-care. For study III, the



data were collected as part of the raters' daily work with their own clients but in the children's own day-care environments.

All the present raters and parents received written and verbal information about the study and about their rights to withdraw and anonymity before participation. Parental consent was requested and signed permissions obtained only after parents had volunteered to participate in the study. Although parents have decision-making power when young children are the object of research, the children should also have the right to decide when they are capable of understanding what consent to being researched means. Hence, the starting point should always be to obtain the parents' informed consent when under-age children are to participate in research and, when possible, also that of the children (see Gallagher, 2009).

In study I and partly in study II, the assessment forms were sent to the day-care centers where the teachers and nurses had expressed their commitment to undertake data collection for the study. Each of the teachers, special teachers, and day-care nurses selected one or more children in his or her daycare group and scored each item based on his or her own observations of the play performance of each child. The observation time was not controlled. It is recommended that the observer and the child had been working in the same group during the previous 2 months. The reason for this was that we wanted the play context to be familiar to the child and the child's play style to be familiar to the observer. In study II some of the raters, and in studies III and IV all of the raters visited day-care groups. They spent approximately 3 to 4 hours per day observing two to three children. Our aim was to ensure that the children should be as relaxed as possible in the observation situations.

Each child was scored individually but in a group situation. It was not required that the child should take part in the group play, but, they had to have possibilities to play with peers if they choose to do so. The raters informed all the children who were participating in play that the rater is interested in seeing how they play. The raters asked the children if they would be allowed to observe what is happening in their play (see Strandell, 2005). The children were encouraged to play as they usually do. If a child asked the rater to play with him or her, the raters encouraged them to continue playing with the other children. Our experience suggests that children are willing to make contact with raters, which we interpret as a display of confidence. All the children seemed to be relaxed in the presence of an observer, and none of them refused the rater's request.

A further ethical consideration is the possible impact of the research on the children, both at the time and at a later date. In addition, participation in the research must always be in the child's best interests. We investigated the play performance of children in a natural play environment. Most of the raters were professionals working in the same day-care center the child was attending. In addition, professionals have a responsibility to ensure that the instruments they use are practically useful and not harmful to children or their families. Therefore, we believe that a measurement situation where the rater already has

a good knowledge of the play profile of a child can only have a positive effect on the child's education, all the more so if a child's play performance gives cause for some concern, in which case it might have extra value in planning interventions with that child.

## 6 PRACTICAL IMPLICATIONS

While play has been seen as a fundamental learning arena for children to acquire skills needed in daily life, the challenges in play performance need to be taken seriously. Despite the fact that one of the core values in the national curriculum guidelines on early childhood education and care is to secure healthy and safe environments that allow play performance for a child (STAKES, 2004), and that play performance is a primary activity of children, no validated play assessment tools are currently in use in the Finnish day-care system. Professionals often make judgments and decisions based on informal and unstructured observations. This means that their observations vary and decision making differs from person to person or day to day. In order to be able to plan individual educational or rehabilitation interventions more efficiently, we definitely need more objective and systematically gathered information on children's play performance and how they approach their environment during play. Only a few valid tools, developed in a different culture and language, and for use by specific professions, e.g. occupational therapists, exist and none of these have yet been implemented in the Finnish day-care system. Consequently, there was a need to develop a structured and usable tool that would efficiently capture the challenges and strengths in a child's observable play performance in daily life environments. I therefore conceptualized the theoretical frame of play performance and constructed the items for the PAGS. It was shown in the studies comprising this thesis that the PAGS can be used as a valid and structured tool to measure a child's play performance in daily social situations, and thus to identify not only the challenges experienced by the child but also the child's strengths. The results indicated that, despite the group-specific play items, the PAGS provides objective information and the necessary structure for play performance observations.

In developing the PAGS, we succeeded in constructing items that not only worked together as an ability continuum for play performance, but also reflected the theoretical background of the PAGS. The continuum of items indicated the lesser or greater play ability of children. Thus, the PAGS was able to separate children with low play ability from those with high play ability. This is critical for any valid and usable assessment tool. The PAGS makes it possible to measure a child's total level of play performance and, more

importantly, identify the specific challenges or strengths of the child's play. If a child shows poor ability at any of the play items that are expected to be relatively easy, this may indicate that this level of play performance should be the primary target of an intervention to improve overall performance. On the other hand, while a child may experience several challenges in other daily situations, in play performance the same child might also show surprising strengths. These strengths in play can be utilized to support efforts to habilitate other challenges. The play profile obtained from the PAGES assessment can be used on the individual level qualitatively to support the reasoning process and decision making of professionals working daily with children. When the rater is familiar with the theoretical elements of the PAGES and understands the basic principles of the MFR approach, h/she is able to interpret the results of the play performance assessment item by item.

One requirement of an assessment tool is that it can be used to indicate changes in the ability it measures. In the evaluation of the test-retest responses of the children's PAGES scorings at a one-week interval, our results revealed the presence of variance in the children's play performance measures. This can pose a threat to the repeated measures reliability of the PAGES, and thus it needs to be re-evaluated with a larger sample. After studying the reasons for this, we concluded that changes in the play environment sometimes, quite dramatically, affected the children's play performance. Therefore, we would emphasize that when the PAGES is used to measure changes in a child's play performance, it is important to ensure that no changes have occurred in the play environment in the interval between the assessments. This means not only the physical environment but, more importantly, social situations.

A rater needs experience to become a consistent rater. Our results demonstrated that professionals with experience of free play observation are easily able to use the PAGES in a reliable way. Although rater severity might vary slightly, calibration is not needed for practical purposes; however we recommend controlling for rater severity when the PAGES is used for research purposes. Besides experience in observing play, raters need a structured manual and enough knowledge on the background theory of the PAGES in order to administer and score the PAGES items in a valid and reliable manner. User education on the PAGES can thus be recommended. In order to increase rater consistency, a thorough understanding of the ability being measured, in this case play performance, should form the core of the training process. However, the function of training is not to force raters into agreement with each other, but rather to train them to be self-consistent. This internal consistency will render the measurement of children's play performance more accurate. However, and most importantly, we recommend that when evaluating play performance, the rater knows the child and his/her play style and that the observations are made in several free play situations before implementing the final scoring of the PAGES.

## YHTEENVETO

Leikki on lapsille tärkeä toiminnan areena. Lapsi oppii leikkiessään taitoja, joita hän tarvitsee arjessaan ja oppimisessaan. Kun lapsella on haasteita kehityksessään, se todennäköisimmin näkyy myös leikkimisen taidoissa. Leikki on tärkein oppimisen foorumi alle kouluikäisen lapsen arjessa ja leikkitaidot ovat edellytyksenä vapaalle ja luovalle oppimiselle ja heijastuvat myöhemmin esim. koulussa ja kotona tapahtuvassa oppimisessa. Kun lapsi ei osaa leikkiä, se vaikuttaa lapsen mahdollisuuksiin osallistua arjessaan. Leikin taitavuuden on todettu heijastavan mm. ajattelun, kielen ja oman toiminnan ohjailun ongelmia. Sen vuoksi leikkitaitojen arviointi tulisi sisältyä päiväkotien rutiineihin.

Lapsen kanssa työskentelevät opettajat, hoitajat ja terapeutit tekevät päätöksiä perustuen vapaaseen ja usein sattumanvaraiseen havainnointiin. Kun havainnoinnista kirjataan strukturoimattomia huomioita, niin päätöksenteko voi vaihdella havainnointitilanteesta toiseen. Tämänhetkisen tietoni mukaan suomalaisessa päivähoitossa ei ole ollut käytössä tutkittua pätevää ja luotettavaa leikinarvioinnin välinettä. Maailmalla olemassa olevat muutamat arviointivälineet ovat kehitetty toisessa kulttuurissa ja kielellä, eikä virallisia suomenkielisiä käännöksiä ja kulttuurista validointia ole tehty. Lisäksi useiden menetelmien käyttö vaatii erillisen menetelmäkoulutuksen ja ne ovat yleensä suunnattu vain tietyn ammattikunnan käyttöön. Uuden helppokäyttöisen, strukturoidun ja validin menetelmän keittämiselle Suomessa oli siis selkeä tarve.

Ryhmätilanteessa tapahtuva lapsen leikin arviointi menetelmä, The Play Assessment in Group Settings (PAGS), tarjoaa nopeasti toteutettavan tutkitun ja validin menetelmän tunnistaa leikkitoiminnassa ilmeneviä lapsen vahvuuksia ja haasteita. Menetelmän avulla lasten kanssa työskentelevät voivat havainnoida ja kirjata tutkittuun tietoon perustuen strukturoidusti lapsen leikkitekojen laatua. Saadun tiedon avulla voidaan paremmin tukea ja ohjata lapsen yksilöllistä kehittymistä ja huomioida lapsen vahvuuksia ja erityishaasteita. Menetelmän kehitystyö alkoi teoreettisen lähestymistavan kuvaamisella ja lapsen leikkitoimintaa määrittävien osoittimien kuvaamisella. Päätavoitteena väitöskirjan tutkimuksissa oli tutkia leikin havainnoinnin välineen, PAGS:n validiteettia ja reliabiliteettia neljässä osatutkimuksessa. Tutkimusaineisto kerättiin havainnoimalla 1 vuotta 8 kuukauden – 8 vuotta 9 kuukauden ikäisiä lapsia päivähoitossa ja esikoulussa. Tutkimusaineiston analyysi tapahtui moni fasettista Raschin mallia (MFR) käyttäen. MFR analyysiä käytettiin tutkimaan menetelmän sisäistä rakennevaliditeettia, validiteettia suhteessa muihin muuttujiin, lasten pisteiden ja arvioitsijoiden pisteytysten Rasch mallin oletuksiin sopivuutta, sekä arviointien pysyvyyttä kahden eri arviointikertojen välillä.

Tutkimuksessa I osoitettiin PAGS:n osoittimien kuvaavan teoreettisesti yksilöteista leikin taitavuutta kuvaavaa leikin jatkumoa. Osoittimet sijoittuvat tälle leikin taitavuuden jatkumolle siten, että ne vaihtelevat helpoista osoittimista aina vaativiin osoittimiin. Tehtyjen havaintojen pohjalta kyetään kuvaamaan lapsen leikin taitavuutta kokonaisluvulla (logit) sekä arvioimaan lapsen leikin laatua yksittäisiä osoittimien pisteytyksiä tarkastelemalla. Kun lapsella on haasteita leikin taitavuudessa, voidaan hänen haasteitaan ja vahvuuksiaan leikissä yksilöidä tarkemmin ja siten hyödyntää tuloksia leikki- interventioiden suunnittelussa.

Vertailtaessa lapsia, joilla on erityisiä kielellisiä haasteita (SLI), normaalisti kehittyviin lapsiin osoitettiin (tutkimus II) merkitsevästi, että kielenkehityksessään jäljessä olevien lasten leikkitaidot olivat alhaisempia kuin ikäistensä verrokkien leikkitaidot. Tutkimuksissa osoitettiin siis selkeästi PAGES:n kyky erotella lapset, joilla on haasteita leikissään normaalisti kehittyvistä lapsista. Lisäksi PAGES:n osoittimista kielellistä kykyä vaativat osoittimet olivat kielenkehityksessään jäljessä oleville lapsille suhteellisesti vielä haasteellisimpia. Päättelimme tästä, että PAGES:ssa on osoittimia, jotka toimivat lapsiryhmästä riippuen eri tavoin. Kieli ja leikki nivoutuvat yhteen, joten leikkiä arvioitaessa on tärkeää sisällyttää leikkitoiminnan arviointiin myös nämä kielen käyttöä vaativat osoittimet. Lisätutkimusta tarvitaan löytyykö eroja leikin jatkumon yhtenevyudessa erilaisissa lapsiryhmissä, esimerkiksi kulttuurien tai eri diagnoosien välillä. Mikäli PAGES:a käytetään tutkimustarkeituuksiin, on tärkeää kontrolloida löytyykö vertailtavien ryhmien väliltä eritavoin toimivia osoittimia ja vaikuttaako niiden pois jättäminen lapsiryhmien välillä todetun leikkitaitavuuden eron merkitsevyyteen.

Eri ammatissa toimivat arvioitsijat (kuntoutus ja päivähoido) kykenivät luotettavasti pisteyttämään lapsen leikkiä PAGES:n avulla (tutkimus III). He toimivat itsenäisen asiantuntijan tavoin tehdessään strukturoituja havainnoja lasten leikistä. Heidän tiukkuustasonsa vaihteli hieman, mikä oli odotettavaakin. Esimerkiksi Kelly-Vance ja Ryals (2007) painottavat, että testimenetelmän koulutuksen avulla voidaan parantaa arvioijien sisäistä yhtenevyyttä, mutta ei vaikuttaa merkitsevästi heidän tiukkuustasoon. PAGES arviointeja tehdessä arvioitsijoiden tiukkuus vaihteli Tutkimuksessa III suhteellisen vähäisesti, mutta tutkimuksessa IV yhden arvioitsijan tiukkuus poikkesi merkittävästi. Kun tutkimuksessa vertaillaan kahden eri arvioijan tekemiä arviointeja toisiinsa, on suositeltavaa kontrolloida myös testaajien välisen tiukkuustason vaikutuksia arviointitulokseen.

Tutkittaessa (tutkimus IV) PAGES arviointien toistettavuutta, viikon välein toteutettujen kahden eri arvioinnin välinen korrelaatio oli yksilötasolla heikohko (ICC). Tämä johtui osin arvioijien kokemattomuudesta sekä hieman suuremmasta tiukkuustason vaihteluista kuin tutkimuksessa III. Arvioijina toimivat toimintaterapian opiskelijat, joilla ei ollut aikaisempaa kokemusta leikin havainnoinnista. Toisaalta arvioijien tekemien lisähuomioiden perusteella leikkitilanteen haasteellisuus vaihtui arviointien välillä. Ympäristön, sekä sosiaalisen ja fyysisen leikkiympäristön vaikutuksia lapsen leikkiin tulee tutkia lisää.

Kaiken kaikkiaan päätelmänä voi todeta, että PAGES toimii luotettavana arviointivälineenä erottelemaan lapset, joilla on leikissään haasteita. Ammattitaitoiset arvioijat osaavat havainnoida leikkiä luotettavasti PAGES:n avulla. PAGES:ssa on osoittimia, jotka ovat suhteellisesti haasteellisempia lapsille, joilla on kielellisiä erityisvaikeuksia. Tästä johtuen lasten arvioinnit poikkeavat MFR mallista hieman odotettua enemmän. Lisäksi ympäristö vaikuttaa arvioinnin tulokseen. Suositeltavaa on, että jatkotutkimuksessa kontrolloidaan arvioijien välinen tiukkuus, mahdollisesti eri tavoin eri ryhmässä toimivat osoittimet sekä leikki ympäristön vaikutukset. Käytännön työvälineenä PAGES toimii hyödyllisenä välineenä erotellessaan lapsen leikissä olevia haasteita ja vahvuuksia.

## REFERENCES

- Altman, I., & Chemers, M. (1980). *Culture and environment*. Monterey, CA: Books/Cole.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *The standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Occupational Therapy Association (AOTA) (2008). Occupational therapy practice framework: Domain and process (2<sup>nd</sup> ed.). *American Journal of Occupational Therapy*, 62, 625-683.
- Andrich, D. (1988). *Rasch models for measurement*. London: SAGE Publications.
- Ayres, A.J. (1972). *Sensory integration and learning disorders*. Los Angeles: Western Psychological Services.
- Bateson G. (1971). The message, "this is play". In: R. E. Herron, & B. Sutton-Smith, (Eds.), *Child's play* (pp. 261-269). New York: Wiley.
- Bledsoe, N., & Shepherd, J. A. (1982). Study of reliability and validity of a preschool play scale. *American Journal of Occupational Therapy*, 36, 783-788.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in human sciences*. New Jersey: LEA.
- Brentnall J., Bundy, A. C., & Kay, F. C. S. (2008). The Effect of the Length of Observation on Test of Playfulness Scores. *Occupational Therapy Journal of Research: Occupation, Participation and Health*, 28 (3), 133-140.
- Brown, T. & Chi-Wen Chien (2010). Occupation-Centered assessment for children. In s. Rogers (Ed.), *Occupational therapy practice for children: A practical guide for occupational therapists* (pp. 135-159). Oxford: UK: Blackwell Science.
- Bundy, A. C. (1993). Assessment of play and leisure: Delineation of the problem. *American Journal of Occupational Therapy*, 47, 217-222.
- Bundy, A. C. (1997). Play and playfulness: What to look for. In L. D. Parham, & L. S. Fazio (Eds.), *Play in occupational therapy for children* (pp. 52-66). St. Louis: Mosby.
- Bundy, A. C. (2001). Measuring play performance. In: Law M., Baum C., & Dunn W. (Eds.), *Measuring occupational performance* (pp. 89-102). Thorofare, NJ: SLACK.
- Bundy, A. C, Nelson, L., Metzger, M., & Bingaman, K. (2001). Validity and reliability of a test of playfulness. *Occupational Therapy Journal of Research*, 21, 276-292.
- Bundy, A. C., & Murray, E. A., (2002). Sensory integration: A Jean Ayres theory revised. In A. C. Bundy, S. J. Lane, & E. A. Murray (Eds.), *Sensory integration, theory and practice* (2nd ed.) (pp. 227-240). Philadelphia: F.A. Davis Company.
- Burke, J. P. (1993). Play: The life role of the infant and young child. In: J. Chase-Smith (Ed.), *Pediatric occupational therapy and early interventions* (pp. 198-222). Boston: Andover Medical Publications.

- Cameron, D., Lelslie, M., Pollock, N., Stevart, D., Toal, C., & Gaik, S. (2001). The clinical utility of the Test of Playfulness. *Canadian Journal of Occupational Therapy, 68*, 2, 104-111.
- Casby, M.W. (2003). Developmental assessment of play: A model for early intervention. *Communication Disorder Quarterly, 24*, 175-183.
- Christiansen, C. H. (1999). Defining lives: Occupational identity: An essay on competence, coherence, and the creation of meaning. *The American Journal of Occupational Therapy, 53*, 547-558.
- Cordier, R., & Bundy, A. C. (2009). Children and playfulness. In: K. Stagnitti & R. Cooper (Eds.), *Play as therapy: Assessment and therapeutic interventions* (pp. 45-58). London: Jessica Kingsley.
- Coster, W., & Khetani, M.A. (2008). Measuring participation of children with disabilities: issues and challenges. *Disability and Rehabilitation, 30*, 639-638.
- Coto S., Fisher A., & Mayberry W. L. (1995). The assessment of motor and process skills applied cross-culturally to Japanese. *American Journal of Occupational Therapy, 21*, 798-806.
- 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.
- Couch, K., Deitz, J., & Kanny, E. (1998). The role of play in pediatric occupational therapy. *American Journal of Occupational Therapy, 52*, 111-117.
- Csikszentmihalyi, M. (1990). *Flow : the psychology of optimal experience*. New York : Harper & Row.
- Decker, J., Dallmeijer, A.J., & Lankhorst, G.J. (2005). Clinimetrics in rehabilitation medicine: Current issues in developing and applying measurement instruments. *Journal of Rehabilitation Medicine, 37*, 193-201.
- Egan, M., & DeLaat, D. (1994). Considering spirituality in occupational therapy practice. *Canadian Journal of Occupational Therapy, 61*, 95-101.
- Eisert D., & Lamorey S. (1996). Play as a window on child development: The relationship between play and other developmental domains. *Early Education and Development, 7*, 221-235.
- Farrant, B. M., Fletcher, J., & Mayberry, M. T. (2006). Specific language impairment, theory of mind, and visual perspective taking: Evidence for simulation theory and developmental role of language. *Child Development, 77*, 1842-1853.
- Ferland, F. (1997). *PLAY, Children with physical disabilities and occupational therapy: the Ludic Model*. Ottawa: University of Ottawa Press.
- Fisher, A. G. (1994). Functional assessment and occupation: Critical issues for occupational therapy. *New Zealand Journal of Occupational Therapy, 45*, 13-19.
- Fisher, A. G. (2006). *AMPS, Assessment of Motor and Process Skills. Volume 1: Development, Standardization, and Administration Manual* (6th ed.). Fort Collins, CO: Three Star Pres.



- Fisher, A. G. (2009) Occupational therapy intervention process model: A model for planning and implementing top-down, client-centered, and occupation-based interventions. Colorado, USA: Three Star Press.
- Fisher, W. (1992). Reliability statistics. *Rasch measurement transactions* (serial online) (cited 2011 December 14); 3, 238. Available at: <http://www.rasch.org/rmt/rmt63i.htm>
- Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. United States of America: John Wiley & Sons.
- Forsyth, K., & Kielhofner, G. (1999). Validity of the assessment of communication and interaction skills. *The British Journal of Occupational Therapy*, 62, 69-74.
- Fujiki, M., Brinton, B., Hart, C. H. & Fitzgerald, A. H. (1999). Peer acceptance and friendship in children with specific language impairment. *Topics in Language Disorders*, 19, 34-48.
- Gallagher, M. (2009). Ethics. In E. K. M. Tisdall, J. M. Davis, & M. Gallagher (Eds.), *Researching with children and young people* (pp. 11-28). London: SAGE.
- Harris, P. L. (1992). From stimulation to folk psychology: The case for development. *Mind and Language*, 7, 120-144.
- Harris, P. L. (1996). Desires beliefs, and language. In P. Carruthers, P.K. Smith, (Eds.), *Theories of theories of mind* (pp. 200-220). Cambridge, UK: Cambridge University Press.
- Harrison, H., & Kielhofner, G. (1986). Examining the reliability and validity of the Preschool Play Scale with handicapped children. *American Journal of Occupational Therapy*, 40, 167-173.
- Hartley, R. E., & Goldenson, R. M. (1963). *The complete book of children's play*. New York: Crowell.
- Hasselkus, B. R. (2002). *The Meaning of Everyday Occupation*. Thorofare: SLACK.
- Haynes, S. N., Richard, D. C., & Kubany, E. S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment*, 7, 238-247.
- Hermansson, L. M., Bodin, L., & Eliasson A-C. (2006). Intra- and interrater reliability of the assessment of capacity for myoelectric control. *Journal of Rehabilitation Medicine*, 38, 118-23.
- Jankovich, M., Mullen, J., Rinear, E., Tanta, K., & Deitz, J. (2008). Revised Knox Preschool Play Scale: Interrater agreement and construct validity. *American Journal of Occupational Therapy*, 62, 2, 221-227.
- Kielhofner, G. (2002). *Model of human occupation: Theory and application* (3rd ed.) Baltimore: Williams & Wilkins.
- Kelly-Vance, L., & Ryalls, B. O., (2007). Best practices in play assessment and intervention. *Best Practices in School Psychology V*, vol. 2, 549-560. (cited 12 December 2011) Available at [http://www.nasponline.org/publications/booksproducts/BP5Samples/549\\_BPV71\\_33.pdf](http://www.nasponline.org/publications/booksproducts/BP5Samples/549_BPV71_33.pdf)

- Kelly-Vance, L., Ryalls, B. O., & Glover, K. G. (2002). The use of play assessment to evaluate the cognitive skills of two- and three-year-old children. *School Psychology International*, 23, 169-85.
- Kelly-Vance, L., & Ryalls, B. O. (2005). A systematic, reliable approach to play assessment in preschoolers. *School Psychology International*, 26, 398-412.
- Knox, S. (1997). Development and current use of the Knox preschool play scale. In: L. D. Parham., & L. S. Fazio (Eds.), *Play in occupational therapy for children* (pp. 35-51). St Louis, MO: Mosby.
- Kooij, R. V. (1989). Research on children's play. *Play & Culture*, 2, 20-34.
- Langer, E. J. (1989). *Mindfulness*. New-York: Addison-Wesley.
- Lautamo, T. (2009). Assessing play in a social setting. In: K. Stagnitti & R. Cooper (Eds.), *Play as therapy: Assessment and therapeutic interventions* (pp. 115-129). London: Jessica Kingsley.
- Lautamo, T., Kottorp, A., & Salminen, A.-L. (2005). Play assessment for group settings: A pilot study to construct an assessment tool. *Scandinavian Journal of Occupational Therapy*, 12, 136-144.
- Law, M. (2002). Participation in occupations of everyday life. *American Journal of Occupational Therapy*, 56, 640-409.
- Law, M., Steinweinder, S., & Leclair, L. (1998). Occupation, Health and Well-being. *Canadian Journal of Occupational Therapy*, 65, 2, 81-91.
- Lawlor, M. C. (2003). The significance of being occupied: The social construction of childhood occupations. *American Journal of Occupational Therapy*, 5, 424-434.
- Learn, J. (1996). The Effect of a Diagnostic Label of Language Delay on Adults' Perceptions of Preschool Children. *Disability*, 29, 582-588.
- Leslie, A. M. (1987). Pretense and representation: The origins of "Theory of Mind". *Psychological Review*, 94, 412-426.
- Lieberman, J. N. (1977). *Playfulness: Its relationship to imagination and creativity*. New York: Academic Press.
- Linacre J. M. (1987-2007). FACETS Many-faceted Rasch measurement computer program. Available at: <http://www.winsteps.com/winsteps.htm>.
- Linacre, J. M. (1989). *Many-faceted Rasch measurement*. Chicago, IL: MESA Press.
- Linacre, J. M. (1994). Sample size and item calibration (or person measure) stability. *Rasch Measurement Transactions*, 7, 328 (cited 12 December 2011). Available at: <http://www.rasch.org/rmt/rmt74m.htm>
- Linacre, J. M. (2005). WINSTEPS Rasch measurement computer program. Available at: <http://www.winsteps.com/winsteps.htm>.
- Linacre, J. M. (2011) *A user's guide to FASETS, Rasch-Model computer programs, program manual 3.68.1* (cited 16 March 2011). Available at: <http://winsteps.com/winman/reliability.htm>.
- Liiva, C. A. & Cleave, P. L. (2005) 'Roles of initiation and responsiveness in access and participation for children with specific language impairment.' *Journal of Speech, Language and Hearing Research*, 48, 868-83.
- Linder, T. W. (1993). *Transdisciplinary play-based assessment: A functional approach to working with young children* (2nd ed). Baltimore: Brookes.

- Linder, T.W. (2008). *The Transdisciplinary Play Based Assessment-2 (TPBA)*. Baltimore: Brookes.
- Mansoor, F., & Houman, B. (2011). The effects of rater training on raters' severity and bias in second language writing assessment. *Iranian Journal of Language Testing*, 1, 1, 1-16.(cited 12 December 2011). Available at: [http://ijlt.ir/portal/index.php?option=com\\_content&view=article&id=3&Itemid=7](http://ijlt.ir/portal/index.php?option=com_content&view=article&id=3&Itemid=7)
- McCabe, P. C. & Marshall, D. J. (2006). Measuring the social competence of preschool children with specific language impairment: Correspondence among informant ratings and behavioural observations. *Topics in Early Special Education*, 26, 234-246.
- Meyer, A. (1977). The philosophy of occupational therapy. *American Journal of Occupational Therapy*, 31, 639-642. (Original work published 1992)
- Nelson, K., Skwerer, K., Goldman, S., Henseler, S., Presler, N., & Walkenfeld, F.F. (2003). Entering a community of minds: An experimental approach to Theory of Mind. *Human Development*, 46, 24-46.
- Neuman, E. A. (1971). *The Elements of Play*. New York: MSS Information.
- Oakley, F., Duran, L., Fisher, A. G., & Merrit, B. (2003). Differences in motor skills in persons with and without Alzheimer's disease. *Australian Occupational Therapy Journal*. 50, 72-78.
- Oxford english dictionary* (2 nd. Ed.) (1989). Oxford: Clarendon Press.
- Pellegrini, A. D., Dupuis, D., & Smith, P. K. (2007). Play in evolution and development. *Developmental review*, 27, 261-276.
- Piaget J. (1962). *Play, dreams, and imitation in childhood*. New York: Norton; 1962.
- Picone, M. & McCabe, P. C. (2005). The reliability and discriminant validity of the Social Interactive Coding System with language impaired preschoolers. *Journal of Early Childhood and Infant Psychology*, 1, 113-128.
- Reilly M. (1974). *Play as exploratory learning*. Beverly Hills, CA: Sage Publications.
- Rescorla, L. & Goossens, M. (1992). Symbolic play development in toddlers with expressive specific language impairment. *Journal of Speech, Language, and Hearing Research*, 35, 1290-1302.
- Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, 57, 221-237.
- Rubin, K., Fein, G. G., & Vandenberg, B. (1983). Play. In P.H. Mussen (Ed.). *Handbook of child psychology: Vol 4, socialization, personality and social development* (4th ed.). New York: Wiley.
- Schuck, P. 2004. Quality Assessing reproducibility for interval data in health-related quality of life questionnaires: Which coefficient should be used? *Quality of Life Research*, 13, 571-586.
- Shutton-Smith, B. (1997). *The ambiguity of play*. Cambridge: Harvard.
- Sick, J. (2011). Rasch measurement and factor analysis. *SHIKEN: JALT Testing and Evaluation SIG Newsletter*, 15,1, 15-17. (cited 2. May 2012). Available at <http://jalt.org/test/PDF/Sick6.pdf>
- Stagnitti, K. (2009). Pretended play assessment. In: K. Stagnitti & R. Cooper (Eds.), *Play as therapy: Assessment and therapeutic interventions* (pp. 87-101). London: Jessica Kingsley.

- Stagnitti, K., Unsworth, C., & Rodger, S. (2000). Development of an assessment to identify play behaviours that discriminate between the play of typical preschoolers and preschoolers with pre-academic problems. *Canadian Journal of Occupational Therapy*, 67, 291-303.
- STAKES, (2004). National curriculum guidelines on early childhood education and care in Finland. (Cited 10.1.2012) Available in <http://www.thl.fi/thl-client/pdfs/267671cb-0ec0-4039-b97b-7ac6ce6b9c10>
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4). PAREonline. net (cited 8 December, 2011) Available at: <http://PAREonline.net/getvn.asp?v=9&n=4>
- Strandell, H., (2005). Lapset, etiikka ja vaikea osallisuus. In M. Heikkilä & M. Pajukoski (eds.), *Laadullisen sosiaalityökimuksen eettiset kysymykset* (pp. 31-38) Helsinki: Stakes.
- Sturgess, J. (2009). Play skill assessment for middle childhood. In: K. Stagnitti & R. Cooper (Eds.), *Play as therapy: Assessment and therapeutic interventions* (pp. 102-114). London: Jessica Kingsley.
- Sumsion, T. & Law, M. A., (2006). A review of evidence on the conceptual elements informing client-centered practice. *Canadian Journal of Occupational Therapy*, 73, 153-162.
- Symons, D. K. (2004). Mental state discourse, theory of mind, and internalization of self-other understanding. *Developmental Review* 24, 159-188.
- Tennant, A., & Pallant, J.F. (2007). DIF matters: A practical approach to test if Differential Item Functioning makes a difference (cited 12 December, 2011). *Rasch Measurement Transactions*, 20, 4, 1082-84. Available at: <http://www.rasch.org/rmt/rmt203e.htm>.
- Tesio, L. (2003). Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *Journal of Rehabilitation Medicine*, 35, 3, 105-115.
- Tham, K., Bernsbång, B., & Fisher, A. G. (1999). Development of Assessment of Awareness of Disability. *Scandinavian Journal of Occupational Therapy*, 4, 184-190.
- Tristan, A. (2006). An Adjustment for sample size in DIF analysis, Tristan A. *Rasch Measurement Transactions*, 20:3 p. 1070-1071 (cited 13 December 2011). Available at: <http://www.rasch.org/rmt/rmt203e.htm>.
- Ueda, S. & Okawa, Y. (2003). The subjective dimension of functioning and disability: What is it and what is it for? *Disability and Rehabilitation*, 25, 596-601.
- Weigle, S. C., (1998). Using FACETS to model rater training effects. *Language Testing*, 15, 263-287.
- Weinstock-Zoltnick, G. & Hinojosa, J. (2004). Bottom-up or top-down evaluation: Is one better than other? *American Journal of Occupational Therapy*, 58, 594-599.
- Wilcock, A. (1993). A Theory of human need for occupation. *Journal of Occupational science*, 1, 17-24.

- Wilcock, A. (1999). Reflections on doing, being, and becoming. *Australian Occupational Therapy Journal*, 46, 1-11.
- World Health Organization (WHO) (2001). International Classification of Functioning and , Disability, and Health, ICF. Geneva, Switzerland: World Health Organization.
- Wright, B. D. (1996). Reliability and separation. *Rasch Measurement Transactions*, 9, 472 (cited 16 March 2009). Available at: <http://www.rasch.org/rmt/rmt94n.htm>.
- Wright, B. D. & Linacre, J. M. (1989). Observations are always Ordinal; Measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, 70, 857-860.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.
- World Health Organization (2001). *International classification of functioning, disability and health*. Geneva, Switzerland: WHO.
- Vygotsky, L.S. (1976). Play and its role in mental development of the child. In J. S. Bruner, A. Joly & K. Sylva (Eds.), *Play*. New York: Basis Books.
- Vygotsky, L. S. (1978). The role of play in development. In: Cole M, John-Steiner V, Scribner S, Souberman E, (Eds.), *Mind in society* (pp. 92-104) Cambridge, MA: Harvard University Press.

## ORIGINAL PAPERS

### I

#### PLAY ASSESSMENT FOR GROUP SETTINGS: A PILOT STUDY TO CONSTRUCT AN ASSESSMENT TOOL

by

Lautamo, T., Kottorp, A., & Salminen, A-L. (2005)

Scandinavian Journal of Occupational Therapy, 12: 136-144

Reproduced with kind permission of the publisher

ORIGINAL ARTICLE

## Play assessment for group settings: A pilot study to construct an assessment tool

TIINA LAUTAMO<sup>1</sup>, ANDERS KOTTORP<sup>2</sup> & ANNA-LIISA SALMINEN<sup>1</sup>

<sup>1</sup>Department of Health Science, Occupational Therapy, University of Jyväskylä, Finland, and <sup>2</sup>Department of Community Medicine and Rehabilitation, Occupational Therapy, Umeå University, Sweden

### Abstract

The Play Assessment for Group Setting (PAGS) was constructed to measure children's play performance. The study was undertaken with 93 children aged from 2 to 8 years to examine whether the items of the PAGS construct a unidimensional scale that can be used to measure children's play ability. The internal scale validity and the person response validity of the PAGS were investigated by examining the goodness-of-fit of the play items and children's play performance to the many-faceted Rasch model for the PAGS. In total, 46 of the 51 play items demonstrated acceptable goodness-of-fit. Of the 93 children, 90.3% demonstrated acceptable goodness-of-fit on the scale of play items. Overall, the results support both the scale and person response validity for the PAGS, as well as providing preliminary evidence that the PAGS can be used for a more detailed evaluation of children's abilities in play performance within natural day-care contexts. Further research is needed to examine other aspects of the validity and reliability of the PAGS measures.

**Key words:** *Play ability, play performance, Rasch analyses, test development*

### Introduction

Play is a phenomenon that has been researched by many disciplines including occupational therapy (1–6). Despite the fact that theoretical play knowledge has increased and the focus of interventions in occupational therapy has changed towards valuing play for its own sake, there are still insufficient and appropriate tools for the assessment of play. It is not an easy task to construct an assessment of play, as play is an elusive concept that is difficult to define (7–9).

Play is an essential part of children's development (10,11). Occupational therapists have been concerned about children's play as an occupational role—how children assume their roles as playmates (12). On the other hand, play has also been viewed as an indicator of other abilities or a child's developmental level (6,13–15). Occupational science, however, shifts the focus away from a functional view and role theory and explores play as an occupation—its features and how it is supported

by intrapersonal and environmental factors. Lawlor (16) argued that we should be shifting the focus of our interests toward the study of children as socially occupied beings. If we want to understand play as a social occupation, we need to study play as social performance, not just as “doing”, but as a matter of “doing with” (see Lawlor (16), p. 426). When evaluating features of play performance, it is important to establish how meaningful, spontaneous, curious, and exploratory activity turns into the mindful, negotiable, and shared social reality of play stories (2,17).

There is a limited number of play assessments reported relative to the amount of literature on play (18). Moreover, Bundy (8) has argued that “the inclusion of the word play in the title of the assessment does not mean that the assessment will provide a complete evaluation of play” (see Bundy (8), p. 90). Play performance is crucial in the child's early development and, therefore, the concept of developmental appropriateness has been extended to the assessment of children's skills and capacities

Correspondence: Tiina Lautamo, OTR, MsC, Sorvaamo 10 a 3, SF-40100, Jyväskylä, Finland. Tel: GSM +358405689206. Fax: +358142602141. E-mail: Tiina.Lautamo@sport.jyu.fi

(Received 17 June 2004; accepted 10 March 2005)

ISSN 1103-8128 print/ISSN 1651-2014 online © 2005 Taylor & Francis  
DOI: 10.1080/11038120510031798



during play (6,8,13–15,18,19). However, it is important to note that many tools are used for play assessment, but that they are used to assess children's internal capacities rather than play per se. (8,14,18).

The Child Initiated Pretended Play Assessment, ChIPPA (13) is a promising tool to measure children's play performance by counting elaborate pretended play actions during play with structured toy settings. The limitation of the ChIPPA is that it was formed to be used in clinical settings and therefore the social context of play is ignored. The ChIPPA is a norm-reference test. An example of a tool used to assess children's play in familiar settings is the Revised Knox Preschool Play Scale (PPS-R) (13). The PPS-R is designed primarily to provide a developmental description of typical behaviour. Validity and reliability studies have only been performed with the earlier version of the PPS-R (20,21). The PPS-R has been reported to be clinically useful in the assessment of children when it is not possible to test them using other standardized tests (8,13).

Bundy (7,8,22) has developed the Test of Playfulness (ToP) using Rasch analyses in order to construct a linear continuum of more or less playful behaviours based on observations in at least two different play situations. The term playfulness was originally defined by Lieberman (5,23). The ToP reflects three elements of playfulness: intrinsic motivation, suspension of reality, and internal locus of control. Bundy also incorporates a fourth aspect of play called framing: giving and reading the cues of other players, as was originally discussed by Bateson (24). The measurement of playfulness enables occupational therapists to focus on the complexity of performance within meaningful activity (7,23). Bundy et al. (22) have provided preliminary evidence of the validity and reliability of the ToP test items by using Rasch analyses. To use the ToP in a reliable way, the user must have special training. Further, it has not yet been possible to derive measures of playfulness from the assessment and the ToP is not yet available for practitioners (22).

The Paediatric Volitional Questionnaire (PVQ) (17,25) is an assessment tool constructed to describe the developmental process of a child's own volition. Reilly (26) identified that play performance develops along a continuum from exploration to competence, and then to achievement. In fact, this sequence is also used in the PVQ. It is a qualitative observational tool with internal validity that has been examined using Rasch analyses (17,25). The PVQ describes children's volitional behaviour and examines motivational issues in everyday situations. It is not a play assessment tool, but there are elements that can be

observed during children's play. The PVQ focuses on the volitional component of behaviour, as well as on the spontaneous behaviour that is also elicited during play.

In summary, the existing play assessments reflect all the different features of play and are used either in clinical settings or in the natural environment. There is a need for a new assessment that reflects occupational therapy theories and the social perspectives of play (8,16,18).

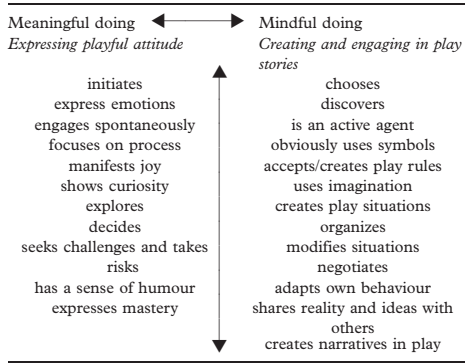
The Play Assessment in Group Setting (PAGS) was designed by the first author of this article to measure the play of 2- to 8-year-old children as performance in the natural day-care context. The main reason for the development of this new instrument was to provide a means to assess children's play performance at an early enough stage to recognize children's occupational challenges in natural social contexts, and so also enable earlier interventions. In addition, the assessment tool must be easy to use by professionals who work daily with children, such as teachers and day-care nurses, and not just therapists. The tool should also be able to be used to measure changes in children's play performance following interventions.

The theoretical features of play performance in the PAGS (see Table I) are based on the concepts of (a) meaningful doing and playful attitude; and (b) mindful doing, observably organized, and socially constructed imaginative play behaviour. These features are combined from theories of playfulness (2,5,7,23) and cognitive psychology theories that reflect on the role of pretended play in cognitive, literacy, and social development (27–31). Consequently, the play features in the PAGS were named as (a) expressing a playful attitude; and (b) creating and engaging in play stories. These features are not viewed as separate or hierarchical. Moreover, they are viewed as dynamically developing and interacting with each other. The 54 items of the PAGS were formed to describe the theoretical features of play performance in everyday language. Items were also formed so that they can indicate less or more play performance. Environmental factors are taken into consideration since the items are scored based on the play performance of children, which is observed in familiar group settings.

When a child expresses a playful attitude, he or she, for example, expresses emotions, shows curiosity, explores, engages spontaneously, and exhibits a sense of humour during play. This feature also describes how children approach social and physical play environments (2,5,7,23,33). When children create and engage in play stories, they use toys, for example, in a symbolic manner and they use their own imaginations to construct play themes. Children



Table I. Theoretical features of play in the PAGS.



create their own ideas, adapt their play performance, as well as adjust their play situations (27,29,31).

Children’s play performances change from object-related reality to symbolically based imagination and role play, and later to more rule-dependent performance (26,28,29). The items of the PAGS were developed to represent this developmental acquisition of play performance. There are easy items, not-so-easy items, more difficult items, and even more difficult items (33). When a child develops to the next level of play performance, he or she still uses old play patterns together with new ones (26). Therefore, we decided to use time-related scoring to describe children’s play performance during free-play situations in the day-care context. Each of the 54 items of the PAGS was scored on a 4-point scale indicating the relative amount of time that a child’s level of “doing” is reflected by that item (1 = hardly ever; the child spends less than 5% of the time engaged in free play, 2 = seldom; the child spends 5–35% of the time engaged in free play, 3 = often, the child spends 35–75% of the time engaged in free play, and 4 = almost always; the child spends more than 75% of the time engaged in free play).

The purpose of this study was to examine whether the items of the PAGS work together to define a single construct that can be used to measure children’s play performance. The following specific research questions were addressed:

1. Do the PAGS items, which are representative of the two theoretical features of play performance, define a single unidimensional construct, such that the items demonstrate acceptable goodness-of-fit with the simple Rasch model?

2. Are the PAGS items targeted to children’s levels of ability in play performance and do the items separate children into different levels of ability?
3. Are the estimates of children’s play performance and the estimates of item difficulties associated with reasonable standard errors?

**Materials and method**

*Participants*

The participants in this study were 93 children aged from 2 to 8 years and who had been in the same day-care group for at least a 2-month period. The participants included children who were developing typically (n = 70) and children whose teachers had either expressed concern that they may have some developmental or learning difficulties or that they had a diagnosed disorder (n = 23). All the diagnoses were based on the reports of the teachers or nurses who worked with the children. The demographic characteristics of the children are described in Table II.

*Procedures*

Prior to commencement of the present study, ethical approval for research was obtained from the manager of Social and Health Services in the city where the children attended day-care. Four occupational therapists and one day-care teacher administrated the pilot version, and after feedback, the tool was modified to create research version 1, which was used to collect the data for this research. The data were collected in nine different day-care centres located in different neighbourhoods in a relatively small town in Finland (80 000 inhabitants). Day-care centres are specially planned for children’s

Table II. Demographic characteristics of the children by groups.

	Typical n = 70 (76.1%)	At risk or disabled n = 22 (23.9%)	Total n = 93 (100%)
Gender			
Male	33	17 (16 dld, 1 a)	50 (53.8%)
Female	37	6 (1 dld, 1 cp, 1 f, 3 r)	43 (46.2%)
Age			
M	4.7	5.6	
SD	1.5	3.2	
Range	2.2–6.9	2.5–8.8	

dld = developmental language learning disabilities, a = Asperger’s syndrome, cp = cerebral palsy, f = fetal alcohol syndrome, r = at risk.

day-care purposes. The children in this study spent approximately 8 hours a day in day-care, and there were several facilities and varieties of toys available daily for free-play situations. The day-care group size varied from 12 to 22 children and there were 3 to 5 teachers or nurses in each day-care group. In three of the nine day-care centres child groups for children with learning difficulties were integrated and special teachers also worked with these children.

Consent for the children to be assessed with the PAGES was provided by their parents. Assessment forms were sent to the day-care centres where the teachers and nurses had expressed their commitment to data collection for this study. Each of the teachers, special teachers, and day-care nurses selected one or more children in his or her day-care group and scored each item based on his or her own observations of each child's play performance during the previous 2-month period. All the raters had professional education in a field of early education and had experience observing and supporting children's play. The raters were not provided with any training in the use of the PAGES except for a short written description on how to score the PAGES items. Six of the children with diagnosed disabilities were assessed by an occupational therapist. The scoring forms without the children's names were returned in envelopes to the researchers.

#### Data analysis

After data collection and careful theoretical reflection, three items that did not necessarily focus towards increasing play ability were removed prior to data analysis (34). Data analysis was implemented using Rasch analyses. Observable performance can only be evaluated and scored using ordinal rating scales. Rating scales are always ordinal and they cannot be treated as quantitative measurements (35). The meaning of scores and estimated quality depends on the specific set of items actually standardized and on the particular ability distribution of the children who happened to appear in the standardized sample (34). Modern statistical test models (e.g. Rasch measurement models) offer an alternative approach to instrument development, which generates unidimensional linear measures based on additive numbers. These models have been described elsewhere in more detail (33–37).

The ordinal raw scores from the scoring forms were entered into a computer and analysed using FACETS, a many-faceted Rasch analyses computer program (37). Two facets were considered in the data analyses of the PAGES: (a) the play performance ability of the children; and (b) the difficulty of the play items. The Rasch model for the PAGES, there-

fore, expected the data to conform to the following two assertions: (a) the child is more likely to obtain higher scores on easier play items than on harder play items; and (b) the easier play items are more likely to be easier for all children than the harder play items. The PAGES play items and children that conform to these expectations will demonstrate acceptable goodness-of-fit with the Rasch model for the PAGES (34).

The FACETS (37) program summarizes all unexpected responses (residuals) into goodness-of-fit statistics. It generates mean square (*MnSq*) and standardized (*z*) goodness-of-fit statistics that were used to evaluate the goodness-of-fit of the children and the play items of the PAGES to the assertions of the Rasch model for the PAGES. The criteria for an acceptable goodness-of-fit (both infit and outfit) of the play items and the person abilities were set at  $MnSq \leq 1.5$  and  $z < 2$  (38). It is generally expected that 95% of the items or persons should meet the criterion that 5% may misfit by chance at  $z < 2$ . More specifically, the analyses made it possible to consider internal scale validity using the goodness-of-fit statistics for the PAGES play items and the person response validity using the play scale goodness-of-fit statistics for each child assessed with the PAGES.

Standard error (SE) estimates for the separation of items and persons were investigated. The size of the SE is influenced by how well the data fit the model assertions and how well targeted the difficulties of items are to the abilities of the people (33). The SE can also be used as a measure of the precision of the generated individual estimates (39). The criterion representing an acceptable standard for persons was  $SE \leq 0.30$  (40). To further confirm the clinical utility of the test, the separation index and reliability for items and persons were investigated. A separation index  $> 2$  and a reliability coefficient  $> 0.80$  was expected (41).

#### Results

The first FACETS analysis with 51 play items revealed that 4 (7.8%) of the 51 items failed to meet the criteria for acceptable goodness-of-fit. We decided to remove the items that failed to demonstrate acceptable goodness-of-fit step-by-step, as long as the separation index of the persons remained stable or improved (see Table III). When item 19 was removed, which was the item that failed the most to demonstrate acceptable goodness-of-fit ( $MnSq = 2.56$ ,  $z = 2.7$ ), the separation of the persons remained similar at 4.91. In the second analysis, when the step-by step removal process was continued, item 5 was removed, which was now the item that failed

to the most demonstrate acceptable goodness-of-fit ( $MnSq = 1.81$ ,  $z = 3.4$ ). The separation index increased from 4.91 to 4.95. On removal of third misfit item (item 27,  $MnSq = 1.68$ ,  $z = 3.9$ ), when the step-by step removal process was continued, the separation index of persons remain stable, it changed from 4.95 to 4.94. Item 27 was the hardest item in the PAGS and, therefore, we decided to retain it in the analysis. Then the fourth misfit item (item 8,  $MnSq = 1.56$ ,  $z = 3.0$ ) was removed. The separation index of the people remained still stable at 4.95. After the removal of the three misfit items, one new item (item 29,  $MnSq = 1.54$ ,  $z = 2.7$ ) appeared to misfit. After theoretical consideration, this item was also removed and the separation index of persons remained stable at 4.95.

#### Internal scale validity

The final analysis revealed that the remaining 47 play items of the PAGS defined one single unidimensional construct such that the 46 (97.9%) items demonstrated acceptable goodness-of-fit with the Rasch model for PAGS. One of the items (2.1%) did not demonstrate acceptable goodness-of-fit (item 27). All the scale items ( $n = 47$ ) were well distributed along a linear scale and targeted to the play skills of children (see Figure I). The item difficulty calibrations and the goodness-of-fit statistics for the PAGS play skill items are given in detail in Table IV.

#### Separation

The separation index of the children's ability measures after the removal of the final item was 4.95 (reliability = 0.96). This indicated the 47 items of the PAGS separated the children into at least five different levels of ability in play performance. There was only one child (a girl aged 4 years and 9 months) who obtained maximum scores and there was no floor effect. The PAGS differentiated the children who were at risk or disabled from typically develop-

ing ones. In all, 18 of the 23 at risk or disabled children were among the lower half of all children on the play ability continuum, while the age mean of the disabled children's group was higher than in the typically developed children's group (see Table I).

#### Person response validity

When examining the goodness-of-fit of the children after item removal, 84 of the 93 (90.3%) children demonstrated acceptable goodness-of-fit on the PAGS. We expected that 5% of the children would misfit by chance based solely on  $z < 2$ . Therefore, we concluded that the overall rate of misfit was slightly higher than expected. The SE of the children's play performance measures was acceptable (SE mean = 0.26).

Since the number of children that failed to demonstrate acceptable goodness-of-fit exceeded the criterion of 5%, a further analysis was undertaken to evaluate whether any systematic patterns were found among those children. Three of the nine children (33.3%) were children categorized as being at risk or disabled. This was only slightly higher than the proportion in the overall group. Five (55.6%) of the nine children were boys and four were girls. This represented an equal proportion within the misfit group, as was the case overall. Six (66.7%) of the nine children were about six years of age (5.10–6.9 years), while the same age-group distribution in the total group was 28.3%.

#### Discussion

The main purpose of this study was to examine whether the items of the PAGS define a unidimensional construct that can be used to measure children's play performance. Overall, our results supported the internal scale validity for the PAGS and, therefore, suggest that the PAGS could be used for the evaluation of children's play performance within day-care contexts.

Table III. Items removed based on outfit  $MnSq$  that failed to fit to the Rasch model for the PAGS.

	Outfit $MnSq$	Step-by-step removal	Person separation index after removal
19. Child plays with conventional toys, e.g. dolls and toy cars	$MnSq = 2.56$ , $z = 2.7$	removed step 1	4.91
5. Child appropriately follows others' play without disturbing and destroying	$MnSq = 1.81$ , $z = 3.4$	removed step 2	4.95
27. Child breaks the rules in ways that bring more fun into play scenarios	$MnSq = 1.68$ , $z = 3.9$	Retained	4.94
8. Child participates skillfully in running, skipping and climbing	$MnSq = 1.56$ , $z = 3.0$	removed step 3	4.95
29. Child appropriately chooses solitary or social play	$MnSq = 1.54$ , $z = 2.7$	removed step 4	4.95

Measures (logits)	Child	Items
	<b>Higher skills</b>	Harder items
5	* * * *	
4	* * **	
3	* * *** * *	
2	** **** ** *****x ***** ***	27, 42
1	*****xx *****x ***x ***xxx ***xx *x	26 43 16 7, 37, 44 6, 10, 12, 13, 14, 35, 41, 52, 53 15, 36
0	**xx * *** **x xx xx *x *x *	11, 24, 47, 54 31, 38, 45, 46, 49 21, 23, 30, 39, 40, 50, 51 4, 20, 25 33, 34, 48 3, 32
-1	*x *x *	22 1 28
-2	x	
	<b>Lower skills</b>	Easier items
	*orx = child	PAGS play items

# = expressing playful attitude  
 ## = creating and engaging in play stories  
 \* = well child  
 x = child with problems

Figure I. Distribution of play items targeted towards the play abilities of children.

All the scale items ( $n = 47$ ) were well distributed along a linear scale and targeted to the play ability of children. They work together to represent children's play performance. Furthermore, the acceptable SE mean related to the children's play ability measures

indicates that the data fit the model's assertions and supports that the difficulties of the items are well targeted to the skills of the subjects. These results indicate that the PAGS has the potential to be developed further as a tool that can be used to detect differences in children's play abilities. It is clinically important that the assessments used are sensitive enough to detect changes in relation to interventions, in order to provide evidence of their effectiveness.

The number of children that demonstrated less than acceptable goodness-of-fit exceeded the set criterion of 5%. Although the proportion was close to what was expected (9.7%), a high number of misfitting participants represents a threat to test validity (33). The limited number of participants in this study ( $n = 93$ ) did not support an in-depth subgroup analysis to investigate whether there was any systematic bias among the children who did not demonstrate acceptable goodness-of-fit. There was no clear pattern regarding gender or functional impairments among the nine children who did not demonstrate acceptable goodness-of-fit in this study. The higher proportion of children among six-year-olds who were in the misfitting group could indicate that the response patterns on the PAGS within this age group may exhibit unexpected variations in play performance. Changes in play performance occur individually. Future studies with more participants in different age groups are needed to provide further evidence of the validity of the PAGS in relation to age.

Another reason for the higher than expected number of children that demonstrated acceptable goodness-of-fit could be that specific raters did not score the items adequately. In any case, there was no clear evidence of rater misinterpretation in the data. This study did not specifically address the examination of the validity of response processes among the raters (intra-rater reliability). Other evidence has suggested that raters may differ in how strict or lenient they are when rating scales are used (33,42), since rater severity has been shown to account for between one- and two-thirds of the variability in the data (43,44). Therefore, it is critical to evaluate the extent to which these factors may impact on the generated measures of the children's play performance. Furthermore, in order to compare children's play performance across time or settings, adjustments must be made for differences in rater severity. Further research with larger samples of children and raters is required to investigate the impact of rater consistency and severity on the generated PAGS measures.

Play is acknowledged to be a complex activity that contributes significantly to all aspects of the devel-

Table IV. Play item difficulty calibrations (logits) and goodness-of-fit to the Rasch model for the PAGS (the items are described in short form, but are written in short sentences in the PAGS).

		Difficulty		Infit		Outfit	
		Calibration	SE	<i>MnSq</i>	<i>z</i>	<i>MnSq</i>	<i>z</i>
Harder items							
27	Breaks for fun	1.84	0.20	1.59	3.60	1.73	4.2
42	Breaks to modify	1.75	0.16	0.99	0.00	1.00	0.0
26	Teases positively	1.25	0.18	1.33	2.1	1.37	2.4
43	Fairytales theme	1.03	0.17	1.10	0.7	1.11	0.7
16	Adventures	0.84	0.17	1.10	0.7	1.07	0.5
2	Risks for challenge	0.72	0.17	1.13	0.9	1.09	0.6
37	Adapts themes	0.50	0.16	0.81	-1.3	0.79	-1.4
7	Joins fluently	0.49	0.16	0.87	-0.9	0.94	-0.3
44	Symbolic objects	0.44	0.17	1.08	0.6	1.05	0.3
12	Negotiates toys	0.41	0.16	0.96	-0.1	0.95	-0.2
35	Creativity objects	0.40	0.16	0.65	-2.7	0.69	-2.2
14	Takes roles	0.39	0.16	0.90	-0.6	0.87	-0.8
41	Negotiates ideas	0.39	0.17	1.06	0.4	1.19	1.2
6	Adapts for challenge	0.36	0.16	0.91	-0.5	0.89	-0.7
52	Narrates play	0.34	0.16	0.78	-1.5	0.77	-1.6
53	Social fluency	0.34	0.16	0.59	-3.3	0.57	-3.4
10	Constructs	0.31	0.16	0.70	-2.2	0.71	-2.1
13	Finds ideas	0.29	0.16	0.84	-1.0	0.80	-1.3
15	Plays home	0.24	0.17	1.06	0.4	1.03	0.2
36	Plays theme or story	0.21	0.16	0.82	-1.3	0.78	-1.4
47	Takes own role	0.03	0.18	1.25	1.6	1.20	1.2
54	Narrates situations	0.03	0.17	1.12	0.8	1.21	1.3
11	Shares toys	-0.01	0.17	1.00	0.0	0.99	0.0
24	Expresses mastery	-0.02	0.16	0.86	-0.9	0.83	-1.1
45	Invisible objects	-0.10	0.18	1.15	1.0	1.06	0.4
31	Expresses emotions	-0.16	0.17	0.90	-0.6	0.83	-1.0
46	Imagines features	-0.16	0.17	0.77	-1.6	0.72	-1.9
49	Asks for help	-0.16	0.17	1.08	0.5	1.02	0.1
38	Accepts other's play acts	-0.21	0.17	0.89	-0.7	0.97	-0.1
50	Accepts help	-0.27	0.17	0.84	-1.1	0.86	-0.8
30	Engages spontaneously	-0.29	0.18	1.14	0.9	1.22	1.3
40	Discovers own ideas	-0.29	0.17	1.05	0.3	0.94	-0.3
21	Varies play objects	-0.30	0.17	0.78	-1.5	0.74	-1.6
39	Accepts adults' play acts	-0.33	0.18	1.12	0.8	1.30	1.7
23	Plays for fun	-0.38	0.17	0.87	-0.8	0.99	0.0
51	Play rule learning	-0.38	0.17	1.00	0.0	0.89	-0.6
25	Decides	-0.49	0.18	1.13	0.8	1.12	0.7
20	Varying places	-0.52	0.17	0.77	-1.6	0.82	-1.0
4	Explores	-0.55	0.17	1.02	0.1	1.04	0.2
48	Is an active agent	-0.70	0.17	0.82	-1.2	0.77	-1.4
33	Is interested in new things	-0.73	0.17	0.89	-0.7	0.83	-0.9
34	Understands humour	-0.73	0.19	1.19	1.2	1.23	1.2
3	Is more organized	-0.83	0.18	0.83	-1.1	0.77	-1.3
32	Focuses on process	-0.86	0.18	0.94	-0.3	0.83	-0.9
22	Expresses curiosity	-1.15	0.18	0.89	-0.6	0.83	-0.8
1	Initiates self	-1.39	0.19	0.99	0.0	0.86	-0.6
28	Enjoys doing	-1.57	0.20	0.78	-1.4	0.89	-0.4
Easier items							

opment of children (2). The complexity of play demands special skills from the persons observing and scoring children's play performance in social and physical environments (2,35). Sutton-Smith (31), for example, indicated that female teachers in particular can easily misinterpret the noisy and aggressive play of boys. Both a teacher's attitude

towards play and the institutional context of day-care affects how play manifests in children's groups. Raters may, therefore, need a more structured introduction to scoring and reflecting on play from an occupational-based perspective, in order to administer and score the PAGS items in a valid and reliable manner.

The PAGS separated the children with low play performance ability from those with high play performance ability. It is critical that a valid assessment of play performance differentiates between persons with limitations in play performance and those without limitations. The results from this study indicate that the PAGS measures are related to other variables (here children at risk or with a disability) and, therefore, provide further evidence of the validity of the PAGS.

One of the benefits of using the Rasch measurement methods in instrument development is that the analyses generate detailed goodness-of-fit statistics that can target potential disruption to the measurement system. In this analysis, five items (5, 8, 19, 27, and 29) failed to finally fit the Rasch measurement model. The degree of fit of a specific item is based on a summary of the unexpected scores (unexpectedly good or unexpectedly poor) among the subjects on this item. For example, in the PAGS, two of these items (5 and 19) were items that appeared to be easy items since almost all children obtained scores of 3 or 4 on these items. Some children, however, obtained lower scores than expected on these items. These children were overall more competent players, but did play more with symbolic toys (toys with no predetermined use) instead of conventional ones (toys with predetermined use, for example dolls or toy cars). These children self-initiated their involvement in play with others, rather than were following the direction of others. Among other misfit items there were no specific response patterns. By removing the misfit items, the PAGS was found to discriminate better between children's abilities in play performance.

The data in this study were collected among a culturally homogeneous population (no children from other ethnic cultures), which could also impact on how the features of play performance are viewed. Play has many universal features, but when trying to describe it as a process of doing, cultural conceptualizations are added into the interpretation and scoring. Therefore, we propose that cultural and environmental issues in the PAGS requires further investigation, for example, by the addition of items that are also more descriptive of the play performance of boys.

While the PAGS will require further revision and research before it is available for use in practice, it shows promise for the fulfilment of the needs of day-care professionals and occupational therapists to have proper tools for the evaluation of play performance. By using the PAGS occupational therapists and other professionals can focus on children's observed performance instead of internal capacities, which are not necessarily predictive of how a child

manages in everyday life situations (16,42). The PAGS focuses on children's play performance in natural social settings and, therefore, provides a better understanding of their occupational challenges. Overall, the results support both the scale and person response validity of the PAGS.

### Acknowledgements

The authors would like to thank all the day-care professionals who committed to this study. They would also like to thank Jyri Manninen (MSc) for checking the language of this manuscript. This study was supported by grants from Eemil Aaltonen foundation and Nordisk Forskeutdanningsakademi (NorFa).

### References

1. Bundy AC. Assessment of play and leisure: Delineation of the problem. *Am J Occup Ther.* 1993;47:217–22.
2. Ferland F. *PLAY, Children with physical disabilities and occupational therapy: the Ludic Model.* Ottawa: University of Ottawa Press; 1997.
3. Hartley RE, Goldenson RM. *The complete book of children's play.* New York: Crowell; 1963.
4. Kooij RV. Research on children's play. *Play & Culture.* 1989; 2:20–34.
5. Lieberman JN. *Playfulness: Its relationship to imagination and creativity.* New York: Academic Press; 1977.
6. Eisert D, Lamorey S. Play as a window on child development: The relationship between play and other developmental domains. *Early Education and Development.* 1996;7:221–35.
7. Bundy AC. Play and playfulness: What to look for. In: Parham LD, Fazio LS, editors. *Play in occupational therapy for children.* St Louis, MO: Mosby; 1997. p. 52–66.
8. Bundy AC. Measuring play performance. In: Law M, Baum C, Dunn W, editors. *Measuring occupational performance.* Thorofare, NJ: SLACK; 2001. p. 89–102.
9. Parham LD, Primeau LA. Play and occupational therapy. In: Parham LD, Fazio LS, editors. *Play in occupational therapy for children.* St Louis, MO: Mosby; 1997. p. 3–21.
10. Piaget J. *Play, dreams, and imitation in childhood.* New York: Norton; 1962.
11. Vygotsky LS. The role of play in development. In: Cole M, John-Steiner V, Scribner S, Soubberman E, editors. *Mind in society.* Cambridge, MA: Harvard University Press; 1978. p. 92–104.
12. Bruke JP. Play: The life role of the infant and young child. In: Chase-Smith J, editor. *Pediatric occupational therapy and early interventions.* Boston: Andover Medical Publications; 1993. p. 198–222.
13. Knox S. Development and current use of the Knox preschool play scale. In: Parham LD, Fazio LS, editors. *Play in occupational therapy for children.* St Louis, MO: Mosby; 1997. p. 35–51.
14. Stagnitti K, Unsworth C, Rodger S. Development of an assessment to identify play behaviours that discriminate between the play of typical preschoolers and preschoolers with pre-academic problems. *Can J Occup Ther.* 2000;67: 291–303.

15. Kelly-Vance L, Ryalls BO, Glover KG. The use of play assessment to evaluate the cognitive skills of two- and three-year-old children. *School Psychology Int.* 2002;23:169–85.
16. Lawlor MC. The significance of being occupied: The social construction of childhood occupations. *Am J Occup Ther.* 2003;57:24–34.
17. Basu S, Kafkes A, Geist R, Kielhofner G. A user's guide to the paediatric volitional questionnaire (Version 2.0). Chicago: University of Illinois; 2002.
18. Sturges JL. Current trends in assessing children's play. *Br J Occup Ther.* 1997;60:410–4.
19. Framer-Dougan V, Kaszuba T. Reliability and validity of play-based observations: Relationship between play behaviour observation system and standardised measures of cognitive and social skills. *Educational Psychology.* 1999;19:429–40.
20. Bledsoe N, Shepherd J. A study of reliability and validity of a preschool play scale. *Am J Occup Ther.* 1982;36:783–8.
21. Harrison H, Kielhofner G. Examining the reliability and validity of the Preschool Play Scale with handicapped children. *Am J Occup Ther.* 1986;40:167–73.
22. Barnett LA. The playful child: Measurement of a disposition to play. *Play and Culture.* 1991;4:51–74.
23. Bundy AC, Nelson L, Metzger M, Bingaman K. Validity and reliability of a test of playfulness. *Occup Ther J Res.* 2001;21:276–92.
24. Bateson G. The message, "this is play". In: Herron RE, Sutton-Smith B, editors. *Child's play.* New York: Wiley; 1971. p. 261–9.
25. Kielhofner G. Model of human occupation: Theory and application, 3rd ed. Baltimore, MD: Williams & Wilkins; 2002.
26. Reilly M. Play as exploratory learning. Beverly Hills, CA: Sage Publications; 1974.
27. Fein GG. Pretended play: An interactive review. *Child Development.* 1981;52:1095–118.
28. Vygotsky LS. Play and its role in mental development of the child. In: Bruner JS, Joly A, Sylva K, editors. *Play.* New York: Basic Books; 1976. p. 537–54.
29. Elkonin DB. Symbolics and its functions in the play of children. In: Herron RE, Sutton-Smith B, editors. *Child's play.* New York: Wiley; 1971. p. 221–30.
30. Pellegrini A. *Applied child study: A developmental approach.* Hillsdale, NJ: Erlbaum; 1991.
31. Sutton-Smith B. *The ambiguity of play.* Cambridge, MA: Harvard University Press; 1997.
32. Cameron D, Leslie M, Teplicky R, Pollock N, Stewart D, Toal C, Gaik S. The clinical utility of the test of playfulness. *Can J Occup Ther.* 2001;68:104–11.
33. Bond TG, Fox CM. *Applying the Rasch model: Fundamental measurement in human sciences.* New Jersey: LEA.
34. Wright BD, Stone MH. *Best test design, Rasch measurement.* Chicago, IL: Mesa Press; 1979.
35. Wright BD, Linacre JM. Observations are always ordinal: Measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989;70:857–60.
36. Fisher AG. The assessment of IADL motor skills: An application of many-faceted Rasch analysis. *Am J Occup Ther.* 1993;47:319–29.
37. Linacre JM. Many-faceted Rasch measurement computer program 1987–2004. Available at: <http://www.winsteps.com>
38. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Measurement Transactions* [serial online] 1994 [cited 12 February 2004];3:370 [1 screen]. Available at: <http://www.rasch.org/rmt/contents.htm>
39. Kottorp A. Occupation-based evaluation and intervention: Validity of the assessment motor and process skills when used with persons with mental retardation. In: Umeå University Medical Dissertations. Umeå: Umeå University; 2003.
40. Tham K, Bernsbång B, Fisher AG. Development of assessment of awareness of disability. *Scand J Occup Ther.* 1999;4:184–90.
41. Fisher W Jr. Reliability statistics. *Rasch measurement transactions* [serial online] 1992 [cited 19 May 2004];3: 237 [1 screen]. Available at: <http://www.rasch.org/rmt/contents.htm>
42. Fisher AG. *AMPS: The assessment of motor and process skills,* 5th ed. Fort Collins, CO: Three Star Press; 2003.
43. Linacre JM. *Many-faceted Rasch measurement,* 2nd ed. Chicago: MESA Press; 1993.
44. McNamara T. *Measuring second language performance.* London: Addison Wesley Longman; 1996.

## II

### **VALIDITY OF THE PLAY ASSESSMENT FOR GROUP SETTINGS: AN EVALUATION OF DIFFERENTIAL ITEM FUNCTIONING BETWEEN CHILDREN WITH SPECIFIC LANGUAGE IMPAIRMENT AND TYPICALLY DEVELOPING PEERS**

by

Lautamo, T., Laakso, M-L., Aro, T., Ahonen, T., & Törmäkangas T. (2011)

Australian Occupational Therapy Journal, 58: 222-230

Reproduced with kind permission of the publisher



Research Article

## Validity of the Play Assessment for Group Settings: An evaluation of differential item functioning between children with specific language impairment and typically developing peers

Tiina Lautamo,<sup>1</sup> Marja-Leena Laakso,<sup>2</sup> Tuija Aro,<sup>3</sup> Timo Ahonen<sup>3</sup> and  
Kari Törmäkangas<sup>4</sup>

<sup>1</sup>Department of Educational Science, University of Jyväskylä and Department of Rehabilitation and Social Sciences, JAMK University of Applied Sciences, Jyväskylä, <sup>2</sup>Department of Educational Science, University of Jyväskylä, Jyväskylä, <sup>3</sup>Department of Psychology, University of Jyväskylä, Jyväskylä, and <sup>4</sup>Finnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, Finland

**Aim:** The purpose of this study was to evaluate the validity evidence based on internal structure of the Play Assessment for Group Settings (PAGS).

**Methods:** The study was conducted in day care centres by comparing observations of the free play performance of two groups of children: those with specific language impairments (SLIs) ( $n = 55$ ) and those with typical language development ( $n = 55$ ). The participating children were 3 to 6.5 years of age. Data were subjected to many-faceted Rasch analyses and differential item functioning analysis was conducted to identify possible group-specific items in the PAGS. The effect of differentially functioning items on mean play performance measurements between two subgroups of children was controlled with two different independent samples  $t$ -tests.

**Results:** According to the results, 80% of the test items were stable across the groups of children. Four items were harder for children with SLI, and clearly required language skills. Two items were relatively speaking easier for children with SLI, and reflected explorative functioning. The differential functioning items did not affect the

ability of the PAGS to separate the children with SLI in mean play performance from their typically developing peers.

**Conclusion:** We concluded that the PAGS is a valid tool for identifying children who have challenges in their play performance and can be used for clinical purposes. However, the slight variability supports the fact that differential item functioning should be controlled in research when using the PAGS for comparison of different subgroups in play performance.

**KEY WORDS** differential item functioning, play performance, Rasch analyses, specific language impairment.

### Introduction

From an occupational therapy perspective, play performance can be understood as a primary activity in which children participate daily by doing something meaningful by themselves, or more typically with someone else (Lawlor, 2003). Difficulties in playing or participating in socially shared situations may give rise to further problems in learning skills needed in daily life (Pellegrini, Dupuis & Smith, 2007). Therefore, the assessment of children's play abilities provides us with essential knowledge on their performance abilities and the interaction between a child and their environment. Assessing children's play abilities should be a fundamental part of professional work in paediatric rehabilitation and education. It is obvious that assessment of play requires valid and reliable instruments that are able to identify children who potentially need interventions to support the participation in play performance.

The Play Assessment for Group Settings (PAGS) was designed to measure the play performance of two- to

Tiina Lautamo Reg OT, LicHSc; Principal Lecturer. Marja-Leena Laakso PhD; Professor. Tuija Aro PhD; Researcher. Timo Ahonen PhD; Professor. Kari Törmäkangas PhD; Researcher.

Correspondence: Tiina Lautamo, Health and Social Studies, Department of Rehabilitation and Social Sciences, JAMK University of Applied Sciences, FI-40100 Jyväskylä, Finland. Email: tiina.lautamo@jamk.fi

Accepted for publication 10 April 2011.

© 2011 The Authors  
Australian Occupational Therapy Journal © 2011 Occupational  
Therapy Australia

eight-year-old children in natural group settings (e.g. day care) (Lautamo & Heikkilä, 2010; Lautamo, Kottorp & Salminen, 2005). The PAGS is an occupation-based assessment instrument. It is intended to be used as a structured tool to observe the play performance of a child to establish their responses to playful occupational challenges and social play opportunities in daily play settings (Lautamo, 2009). In previous studies, the many-faceted Rasch model (MFR) was used to investigate the validity and reliability of the PAGS. The results supported internal scale validity (item response processes), separation ability (Lautamo *et al.*), and inter-rater reliability (rater response processes) (Lautamo & Heikkilä) of the PAGS. However, the person response validity of the instrument has not been demonstrated to be adequate. In fact, the previous studies have shown that data from almost 10% of the children failed to conform to the expectations of the Rasch model (Lautamo *et al.*). The possible reason for this person response bias might be differential item functioning (DIF) caused, for example, by specific sample characteristics such as gender, age, play style or differential diagnoses (Decker, Dallmeijer & Lankhorst, 2005). Therefore, further research is needed to establish the possible causes of this person response bias of the PAGS.

When estimating a scored phenomenon, in this case play performance, with a Rasch model, all the test items are expected to form a unidimensional continuum representing the theoretical construct of the phenomenon being measured (Bond & Fox, 2007; Tesio, 2003). The existence of this unidimensional continuum of PAGS items was confirmed in an earlier study (Lautamo *et al.*, 2005) when using one heterogeneous group of children and in that study PAGS was considered to reflect reasonable internal scale validity. In addition, the test items are expected to remain stable, but shift upward or downward in proportion to the overall mean difference in values across the different groups studied. The bias found in our previous studies on the person response validity of the PAGS implies that this requirement is possibly not fulfilled. Lack of measurement uniformity may lead to incorrect estimates of effects in research and decision making at the individual level (Decker *et al.*, 2005). In the present study, DIF analyses were used to assess whether the PAGS meets the required test criteria of uniformity and stability and whether it is valid as a generic tool to assess differences between various subgroups (Decker *et al.*).

Differential item functioning occurs when people from different groups have a different probability of getting a particular score on test items. DIF analysis provides an indication of unexpected behaviour by item on a test. Person response validity, therefore, can be examined with DIF analyses using a sample consisting of subgroups differing in essential characteristics that would affect the measured phenomenon. Accordingly, when one or more item parameters differ across groups, an item displays DIF.

In the present study, children with specific language impairment (SLI) were considered as a potential comparison group for typically developing children (TD), as they have been found to show less than optimal development in play performance (Casby, 1997; Rescorla & Goossens, 1992). We expected that if DIF occurred when children with SLI and the TD children were combined, this could be caused by those PAGS items which clearly depend on adequate language skills (e.g. describing what is going on while playing, or discussing the rules of play with playmates). Hence, the bias found in previous studies (Lautamo *et al.*, 2005) could possibly be understood as a reflection of diagnostic related failure to conform to the expectations of the Rasch model. The complex connections between language and play are briefly described, to facilitate the discussion of the potential group specificity of the PAGS items.

The decreased play ability of children with SLI has been convincingly demonstrated. Problems in play performance have been detected both in their individual play abilities and in their collaboration with age mates. According to Casby (1997) and Rescorla and Goossens (1992), children with SLI have been found to spend less time on symbolic play, score lower in symbolic play tests and exhibit less sophisticated spontaneous play, compared to TD children of the same age. Children with SLI have been found to address their play partners significantly less, participate less in group play, and engage more in individual play and onlooking behaviour (Liiva & Cleave, 2005). Overall, their play-related difficulties are most evident in the context where collaboration with peers is required (Fujiki, Brinton, Hart & Fitzgerald, 1999; Holmes, 2002; Miller, 2001).

Participation in social play performance requires the ability to share play reality and to collaborate, the ability to 'read the minds' of other individuals, and 'see' things from another person's perspective (Farrant, Fletcher & Mayberry, 2006). This requires the ability to evaluate actions on the basis of one's own or others' intentions, beliefs and plans (Nelson *et al.*, 2003). According to Harris (1996), both of these socio-cognitive abilities require a process of setting aside your current point of view and sharing another person's perspective in your imagination. Language profoundly facilitates these abilities, as conversation and complex use of language – with repairs and clarifications – involve a constant exchange of differing points of view. Furthermore, McCabe & Marshall, 2006; (Picone & McCabe, 2005) showed that children with SLI express delays in controlling and adjusting their behaviour in the play context, exhibiting, for example, poor task orientation, high ratings of dependency (lack of initiatives), poor frustration tolerance and lack of assertiveness. Therefore, children with SLI rely much more on adults than on their TD peers to help them navigate their social environments.

The main aim of the present study was to examine the validity evidence based on the internal structure of the

PAGS. On the basis of the earlier studies it was expected that children with SLI would generally show more challenges in play performance in social settings compared to TD peers when measured with the PAGS. However, because some of the PAGS items clearly require adequate language skills (e.g. use of symbols and imagination and ability to join in play narratives and play realities created by other children), the specific focus of the present study was to examine differentially functioning items of the PAGS between two groups of children (SLI and TD).

The specific research questions were as follows:

1. Do the items of the PAGS function differentially in the two groups of children: the children with SLI and their gender- and age-matched TD peers?
2. If differentially functioning items are found, do they affect the ability of the PAGS to statistically differentiate the children with SLI and TD peers in play performance?

## Method

### Participants

Participants were selected from a total sample ( $n = 208$ ) of all the children having scores in the database for the PAGS, who met the following criteria: (i) age between three and seven years; and (ii) a diagnosis of SLI or typical development (TD). One hundred and 56 children met these criteria. Seventy eight were TD children and 78 were children with SLI. Diagnosis of SLI was based on parental report. Typical children were reported to have no specific problems in everyday life.

To control for the effects of age and gender on play performance, proportional matching was conducted. The participants were individually matched according to gender, age (within three months), and diagnosis (SLI and TD). Because the SLI group contained more boys and older children, children were removed randomly. The final age- and gender-matched sample comprised of 110 children aged 3.1–6.5 years ( $M = 5.4$ ,  $SD = 0.87$ ): 55 children with SLI (boys  $n = 32$ , girls  $n = 23$ ) and 55 TD children (boys  $n = 32$ , girls  $n = 23$ ).

### Instrumentation

The PAGS is an observation-based instrument designed to be used by professionals working daily with children in group settings (e.g. day care). In developing the PAGS, the MRF model was used to calibrate two facets (items, children) (Lautamo *et al.*, 2005), and for rater reliability purposes, three facets (items, children and raters) (Lautamo & Heikkilä, 2010). MFR analysis was previously used to create the PAGS play scale, which represents the acquisition of play performance. That is, the items range from less challenging to challenging and even more challenging items. The MFR model is an acceptable and useful method for creating an objective measurement in the psychological and social sciences. Each child's ability,

item difficulty and rater severity distribution can be set on the same continuum. These ability measures are comparable with each other as logit values (Bond & Fox, 2007).

The earlier study (Lautamo *et al.*, 2005) confirmed that play performance seems to be a unidimensional phenomenon which can be quantified as a total (logit) measure of play. The PAGS play performance scale has been shown to demonstrate high internal scale validity and high inter-rater reliability and, as noted earlier, the PAGS was able to separate children with low play performance from those with high play performance (Lautamo & Heikkilä, 2010; Lautamo *et al.*). However, in a previous study (Lautamo & Heikkilä, 2010), rater severity differed slightly more than expected ( $\pm 0.51$ ).

When evaluating play performance, the rater used time-related scoring to describe the approximate amount of time the child's play behaviour reflected a certain play item during a free play situation in a natural group setting. Each of the 38 items (presented in Table 1) of the PAGS were scored on a four-point scale, indicating the relative amount of time that a child's play behaviour reflected that item (1, *hardly ever*, less than 5% of the time; 2, *seldom*, 5–35% of the time; 3, *often*, 35–75% of the time; and 4, *almost always*, more than 75% of the time). The time-related scores are approximations of the child's play performance in different play situations. It is recommended that the PAGS play observations be made in several situations.

### Procedures

Prior to its commencement, ethical approval for this study was obtained from the manager of the Social and Health Services in the towns or communities where the children attended day care and also from the ethical committee of the Central Hospital in the County of Central Finland. Furthermore, all the parents gave their written consent to assess the play performance of their children.

All children in the PAGS database were assessed by observing children in day care centres in central Finland, both in urban and rural areas, by a teacher, special teacher or day care nurse, or by an occupational therapist or physiotherapist. All the raters had a professional education in early childhood education or paediatrics and had experience in observing and supporting children's play. The raters were provided with either a short training session with video observations or no training, but with written instructions on how to score the PAGS items. The play observations were made in live free play situations, where it was possible for the child to join in the play of other children. The rater freely chose the observation time and occasion. The raters were encouraged to spend as much time as they needed and observe in different play situations to gain an overall picture of each child's play behaviour.

Day care centres in Finland operate in environments specifically planned for child day care purposes. Chil-

**TABLE 1:** *The Items of the Play Assessment for Group Settings (PAGS)* (Lautamo, 2009; in Stagnitti and Cooper *Play as Therapy: Assessment and Therapeutic Interventions*, Jessica Kingsley Publishers, London and Philadelphia. Reproduced with kind permission of Jessica Kingsley Publisher).

---

The items of the PAGS

---

1. Enjoys activity and play; has fun playing
2. Expresses curiosity towards the environment and the objects
3. Begins to play without adult support
4. Devotes themselves in the process of playing and enjoys the activity
5. Explores the opportunities provided by the surroundings
6. Participates even in new games with curiosity
7. Plays tricks and makes jokes
8. Decides on his or her own actions independently rather than imitating others
9. Takes part in shared play activity as an active participant
10. Embraces activity in new surroundings
11. Expresses feelings during play
12. Expresses the feeling of competence by, for instance, describing accomplishments or expressing contentment
13. Adapts actions to make the play more challenging or enjoyable
14. Takes risks and seeks challenges
15. Teases in a positive and playful manner
16. Playfully breaks the rules of the activity and looks for challenges in order to accommodate the play
17. Plays in an organised manner with an obvious goal
18. Accepts other child's play acts as part of shared play
19. Attributes qualities to objects, e.g. the doll is ill, the car is broken
20. Comes up with their own ideas for playing
21. Uses diverse and varying toys and objects while playing
22. Replaces missing objects with imaginary ones, e.g. pretends there is food in the pot
23. Happily shares toys with other players
24. Learns/understands new play rules
25. Builds play surroundings like houses or huts
26. Play-acts a theme or a story
27. Invents play about everyday situations like being at home or in a shop
28. Names a role for themselves, pretends to be someone else
29. Uses objects symbolically, e.g. a box as a table, a block as a car
30. Uses objects in a creative and original way
31. Joins in an ongoing play adapting to its rules
32. Invents adventure plays like being a princess, a pirate or a spaceman

---

**TABLE 1:** *(Continued)*

---

The items of the PAGS

---

33. Describes what is going on while playing
34. Actively modifies or adapts the theme as the play progresses
35. Discusses the rules of play with other playmates
36. Moves easily between stages of play in collaborative play
37. Invents new play activities and shares ideas with playmates
38. Understands the play rules others have set

---

dren commonly spend approximately eight hours a day in day care, and have several opportunities for free play with the varieties of toys available. Group sizes in day care can vary, but generally there is approximately one teacher or nurse per six children in each group. The children with SLI were all integrated with TD children.

**Data analysis**

Data analysis was implemented using MFR analyses. Observable performance has been evaluated and scored using ordinal rating scales. Rasch measurement models, like some other modern statistical models, offer an alternative approach to instrument development, which generates unidimensional linear measures based on additive numbers. These models have been described elsewhere in more detail (Bond & Fox, 2007).

The ordinal raw scores of children's play performance were analysed by using FACETS (www.winsteps.com), version 3.54.1, a MFR analysis computer program (Linacre, 1987-2008). We conducted two different analyses to transform the children's play performance scores to be able to investigate differences between groups and to determine whether DIF influenced the group differences in mean play performance. For this purpose we used the total data of all children in the database ( $n = 208$ ) of the PAGS to confirm the heterogeneity of the data (Bond & Fox, 2007). In the first transformation analyses, we used the latest version (version 3) of the PAGS with all 38 items (Lautamo, 2009) and in the second transformation analyses we only used uniformly functioning items ( $n = 28$ ) that fitted to the Rasch model ( $MnSq \leq 1.3$ ,  $z < 2$ ). Measurable interval values (logit) of play performance for each child were generated via these logistic transformations. Logistic transformation makes it possible to position each child along the linear continua of play performance. Higher scores indicate a higher level of play performance and lower scores indicate lower level of play performance. Likewise, higher logit values represent more challenging play items and lower logit values represent less challenging play items.

While converting the children's play performance scores via the first logistic transformation, we examined

at the same time the scale validity according to the principles reported in a previous study (Lautamo *et al.*, 2005) but with more strict criteria ( $MnSq \leq 1.3$ ,  $z < 2$ ). Two of the items (items 15 and 18) were found to misfit based on this criterion. These items which did not sufficiently fit the model required further investigation. Misfit items might be deleted to improve the scale (Bond & Fox, 2007). Anyhow, it is recommended (Hagquist & Andrich, 2004) that the potential reasons, such as DIF, to be investigated before items are deleted. Therefore, preliminary DIF analysis was conducted. No meaningful differences (Mantel-Haenszel,  $P < 0.05$ ) in item functioning between the two groups of children on these items were found. Therefore, we considered other more theoretical or practical reasons for misfit. The two misfit items seem to be sensitive on environmental effects. In day care centres in Finland the common behavioural rules might have affected the interpretation of these items (15 and 18). In addition, removing these items did not affect other items fit statistics. Therefore, these two items were removed from the data before the further analyses. In addition, one item (item 38) with very few ratings was also removed. The further DIF analysis was completed on 35 items.

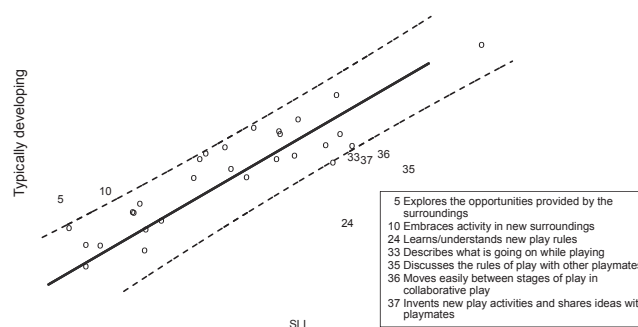
To compare the item functioning of two subgroups: (i) TD children ( $n = 55$ ); and (ii) children with SLI ( $n = 55$ ), DIF analysis was performed to generate group-specific play item difficulty calibrations. This specific MFR analysis was driven using WINSTEPS (www.winsteps.com), version 3.63.0 (Linacre, 2005). To answer the question, whether there were play items that were relatively easier or more difficult for the children with SLI than for the TD children, the item calibrations were plotted with 95% confidence intervals (CI). Furthermore, the DIF probability (Mantel-Haenszel) for meaningful differences between the two groups was set at  $< 0.05$ . Our basic assumption was that the play item difficulty calibration hierarchy would remain stable across different groups. Generally, it was expected that 95% of the items would meet this crite-

riterion, that is, one or two items would lie outside the CI (Bond & Fox, 2007).

To investigate if differentially functioning items and misfitting two items affect the ability of the PAGS to differentiate the children who have challenges in play performance, two different independent samples *t*-tests were conducted. Firstly, using the play performance values (logits) of individual children, conducted with all 38 play items (the PAGS, Current version 3) (Lautamo, 2009), and secondly, with the performance values (logits) of individual children, conducted with the limited set (28) of uniform and stable PAGS items. The effect sizes for the two independent groups were also calculated (Cohen's *d*) (Rosnow & Rosenthal, 2003). Equalities of variance were examined using Levene's test.

## Results

When analysing the DIF, 28 (80%) of the 35 items seemed to be stable across both groups of children. However, seven (20%) items differentiated in challenge between the SLI and TD groups. This was more than expected. Even though the item hierarchy in the PAGS was mainly stable, group-specific play items emerged (Fig. 1). Two items (item five: explores the opportunities provided by the surroundings; and item 10: embraces activity in new surroundings) were relatively less challenging for the children with SLI than for their TD peers. Five items (item 24: learns/understands new play rules; item 33: describes what is going on while playing; item 35: discusses the rules of play with other playmates; item 36: moves easily between stages of play; item 37: invents new play activities and shares ideas with playmates) were relatively more challenging items for children with SLI than their TD peers. The play performance scores of the children and the relative difficulty hierarchies of all the in-fitting items for the two groups of children are presented in Figure 2.



**FIGURE 1:** Differential item functioning plots of item difficulty (in logits) of children with SLI and typically developing children. Items outside the control lines (95% CI) indicate differential item functioning between the two groups.

For the purpose of investigating whether the seven differentially functioning items and the three earlier removed items affect the PAGES ability to differentiate

Measures (logits)	Child	PAGES play items, TD	PAGES play items, SLI
	Higher skills	Harder items	Harder items
+6	xx		
+5	x		
+4	x xxx		
+3	x x xxxx xx x xx		16
+2	xxx x xxxxx xxx xxxxxx xx	16	<b>35, 36, 37</b> <b>33</b> <b>24, 34</b> 13, 14, 30 31
+1	xxxx xxxxxxxx xxxx xxxxxx xxxxxx xxx	14	22, 26, 28, 29, 32 27 20, 23 21, 25 9, 19
0	xxx xxxx x xxx xxx xxxx	29 27, 31, 32 23, 30 25, 28, 34, <b>36</b> 13, 21, <b>33, 35, 37</b> <b>5, 10, 19, 22</b> 11, 12, 20, 26	6, 7, 11 <b>10, 17</b> 4 8 <b>3, 5</b>
-1	xxxxx x xx x x x x xx	6, 9 2, 8, <b>24</b> 4 7, 17 3	2 1
-2	x xx		
-3	x	1	
	Lower skills	Easier items	Easier items

**FIGURE 2:** Targeting of the play abilities of children: Children's play measures in relation to item difficulty calibrations of the two groups of children: typically developing and SLI (differentially functioning items in bold).

children who have challenges in play performance, the independent samples *t*-tests were used. The results revealed significant differences between the two groups of 3.1 to 6.5-year-old children (with and without SLI). In the first analysis of differences in play performance (conducted with 38 items) independent samples *t*-tests confirmed that the means differed significantly ((108),  $t = 5.80, P < 0.01$ ), and the effect size was large (Cohen's  $d = 1.11$ ). In the second analysis of the differences in mean play performance (conducted with 28 stable items), independent samples *t*-test confirmed that the means still differed significantly ((108),  $t = 5.02, P < 0.01$ ), and the effect size was also large (Cohen's  $d = 0.97$ ). Levene's test confirmed the equality of variances in both analyses. Based on these two analyses, we concluded that despite the differentially functioning seven items and two misfitting items, the PAGES (current version 3) demonstrates the power to indicate poor play performance of children with SLI and separate them from their TD peers (Table 2).

**Discussion**

The purpose of this study was to evaluate the validity evidence based on internal structure of the PAGES by investigating the DIF across the two groups of children (SLI and TD). DIF analyses were used to assess whether the PAGES meets the required test criteria of unidimensionality and stability, and whether it is valid as a generic tool to assess differences between various subgroups (see Decker *et al.*, 2005). The results indicated that most of the play items remained stable within the 95% CI. However, the differentially functioning seven items revealed that the PAGES was somewhat unstable across the two groups of children (SLI and TD) because the position of these items differed slightly in the item hierarchy between the two groups. In detailed analysis, we found five items that were relatively more challenging and two items that were relatively less challenging for children with SLI than their TD peers.

**TABLE 2:** Mean difference in play performance of the children with SLI ( $n = 55$ ) and TD peers ( $n = 55$ ) measured with the PAGES: current version 3 with 38 items and limited 28 items without differentially functioning and misfitting items

Participant group	With 38 items $t = 5.80, P < 0.01$		With 28 items $t = 5.02, P < 0.01$	
	Play ability M	SD	Play ability M	SD
TD	1.66	1.55	1.85	1.97
SLI	0.10	1.25	0.20	1.39

TD, typically developing; SLI, specific language impairment.

To understand the possible sources of the different functioning of the items, we need to further consider whether the content of these seven items was consistent with the previously reported problems in play performance of the children with SLI. In our opinion, four of the five relatively more challenging items clearly required expressive or receptive language skills. Of these items, three required the adequate use of expressive language and the fourth one required receptive language skills. The expressive language related items were as follows: item 33 (describes what is going on while playing), item 35 (discusses the rules of play with other playmates), and item 37 (invents new play activities and shares ideas with playmates), and the receptive language related item was item 24 (learns/understands new play rules). In other words, all these four items refer to the difficulties the children with SLI experienced in inventing new ideas and, in particular, sharing them with others by describing what is going on during play.

These considerations are in accordance with previous studies, which focussed on children's participation in peer relationships and emphasised the role of language in play within social settings (Fujiki *et al.*, 1999; Liiva & Cleave, 2005). However, the problems detected in receptive language skills among children with SLI have affected children's play performance (e.g. the time needed to understand what is happening in the play and the ability to participate in play interaction) to a lesser extent than have problems in expressive language (Liiva & Cleave). Consequently, the question arises as to whether language abilities can be separated from the play performance measure. DeKroon, Kyte and Johnson (2002) emphasises that when assessing children's play we should take into account not only aspects of play performance, but also aspects of conversation with peers. Therefore, these items are interpreted to be an important part of the play performance measured with the PAGS and should not be rejected from the tool.

The fifth relatively more challenging item, item 36 (moves easily between stages of play in collaborative play) describes the child's capability to understand and to be flexible in the ever-changing social reality of play. Earlier studies showed that children with SLI are delayed in controlling and adjusting their behaviour (McCabe & Marshall, 2006; Picone & McCabe, 2005) and have difficulties in 'reading the minds' of other individuals, and 'seeing' things from other person's perspective (Farrant *et al.*, 2006). This kind of play performance requires not only language skills, but also the skills to plan and to adapt to the process of doing. We concluded that this item also clearly reflects the typical problems in play in children with SLI reported in earlier research. Hence, it should be included in the PAGS, as it illustrates important play performance, but requires further investigation with other subgroups of children.

Our findings also indicated that there were two items of the PAGS that were relatively less challenging for chil-

dren with SLI than their TD peers. These were item five (explores the opportunities provided by the surroundings), and item 10 (embraces activity in new surroundings). Children learn through exploration of their environment and this kind of behaviour dominates during infancy (Pellegrini *et al.*, 2007). However, this form of play performance is also an important part of play, especially in new environments and situations and reflects the style the child adopts to approach their environment. Anyhow, based on Rasch model assumptions easy items should be easy for all children and our participants were already 3 to 6.5-years old and observations were conducted in familiar environments. These two less challenging items indicated that children with SLI spend relatively more time in activities by exploring and embracing in environment compared their total play performance level. Based on our clinical experience and earlier studies, domination of the exploring activity might reflect their inability to play at a more symbolic level (Rescorla & Goossens, 1992). On the other hand, the deficits in language skills and inability to join fluently to social play (Fujiki *et al.*, 1999) might encourage the child to make the most and they spend more time of play in the surroundings. These results indicate that children with SLI might benefit the support how to join other children's play to be able to take in use capacities they have and to play more sophisticated levels than they usually do.

Because the lack of uniformity in measurement may lead to incorrect estimates of effects in research (see Decker *et al.*, 2005) we provided a control to see whether omitting the differentially functioning items affected the mean difference between two subgroups, children with SLI and their TD peers. In the first analysis, the play performance measure calibrations of the children were conducted with the total set of 38 items (current version 3) of the PAGS (Lautamo, 2009), and in the second analysis we used the calibration data of children with the 28 uniformly well functioning items. No notable differences were found between the results of these two analyses. However, it was important to investigate whether the differentially functioning items had any impact on the significance of the mean difference when the play performance differences of the groups were relevant (see Bond & Fox, 2007). Our results indicated that despite the slight lack of uniformity of the PAGS, it succeeded in differentiating these two groups of children in play performance and therefore can be regarded as a valid tool for research purposes. However, the detected slight variability on the uniformity of the instrument suggests that the DIF be controlled in future studies when comparing play performance of different subgroups such as gender, age, play-style or differential diagnoses of children.

The main limitation of the study was that only the internal aspect of validity was controlled. External aspects, such as rater severity or inter-rater reliability may also affect the results. Because we used several raters to code the data, each of whom made only one or two

observations, it was not possible to control the rater severity. In addition, the raters had different levels of training for scoring the PAGS which might be a risk to the inter-rater reliability. Nevertheless, in an earlier study (Lautamo & Heikkilä, 2010) the results indicated that the PAGS scorings between the diverse professional raters were free from bias and comparable with each other. The results supported that the PAGS can be used as originally intended by all professionals (teachers, day-care nurses and therapists) who work on a daily basis with children.

All the raters in the present study were working daily with the children they scored. Consequently, the raters were aware of the diagnoses of the children which might cause a risk of a tendency to score lower the children with SLI on items that involve language. In addition, the raters were allowed to freely choose the observation time which might have also effected on their judgments. According to Brentnall, Bundy and Kay (2008) when scoring children's playfulness with the Test of Playfulness, an observation-based rating scale tool, the scores based on 30-minute observations were significantly lower than those of either of first or last 15-minute period of observations. Anyhow, the lengthening of the time did not provide additional information nor increase the test-retest reliability. Authors suggested that longer observing time is not necessarily more representative than the shorter time. If the results of Brentnall *et al.* (2008) are replicable with other observation-based assessments, such as PAGS, that raises the challenge of future studies to examine whether the length of observation time effects the observations of the PAGS.

Professionals often make the judgments and decisions based on informal and unstructured observations. Structured tools are needed to more efficiently capture and identify strengths and challenges in child's play performance in social situations. The results of the present study support further the validity evidence based on internal structure of the PAGS and despite the group-specific play items we suggest that PAGS can be used to provide the needed structure for play performance observations.

## Acknowledgements

The authors gratefully acknowledge the professionals, children and their parents who contributed to the data collection for PAGS database. The authors would also like to thank Tarja Karjalainen and Michael Freeman for reviewing the language of this manuscript.

## References

- Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch model: Fundamental measurement in human sciences*. Mahwah: Lawrence Erlbaum.
- Brentnall, J., Bundy, A.C. & Kay, F.C.S. (2008). The effect of the length of observation on Test of Playfulness scores. *Occupation, Participation and Health*, 28 (3), 133–140.
- Casby, M.W. (1997). Symbolic play of children with language impairment: A critical review. *Journal of Speech, Language and Hearing Research*, 40, 468–480.
- Decker, J., Dallmeijer, A.J. & Lankhorst, G.J. (2005). Clinimetrics in rehabilitation medicine: Current issues in developing and applying measurement instruments. *Journal of Rehabilitation Medicine*, 37, 193–201.
- DeKroon, D.M., Kyte, C.S. & Johnson, C.J. (2002). Partner influences on the social pretended play of children with language impairments. *Language, Speech, and Hearing Services in Schools*, 33, 253–267.
- Farrant, B.M., Fletcher, J. & Mayberry, M.T. (2006). Specific language impairment, theory of mind, and visual perspective taking: Evidence for simulation theory and developmental role of language. *Child Development*, 77, 1842–1853.
- Fujiki, M., Brinton, B., Hart, C.H. & Fitzgerald, A.H. (1999). Peer acceptance and friendship in children with specific language impairment. *Topics in Language Disorders*, 19, 34–48.
- Hagquist, C. & Andrich, D. (2004). Is the sense of coherence instrument applicable on adolescents? A latent trait analysis using Rasch modelling. *Personality and Individual Differences*, 36, 955–968.
- Harris, P.L. (1996). Desires beliefs, and language. In: P. Carruthers & P.K. Smith (Eds.), *Theories of theories of mind* (pp. 200–220). Cambridge, UK: Cambridge University Press.
- Holmes, A.M. (2002). Theory of mind and behaviour disorders in children with specific language impairment [Abstract]. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 62, 11.
- Lautamo, T. (2009). Assessing play in a social setting. In: K. Stagnitti & R. Cooper (Eds.), *Play as therapy: Assessment and therapeutic interventions* (pp. 115–129). London: Jessica Kingsley.
- Lautamo, T. & Heikkilä, M. (2010). Inter-rater reliability of the Play Assessment for Group Settings. *Scandinavian Journal of Occupational Therapy*, 18, 3–10.
- Lautamo, T., Kottorp, A. & Salminen, A.-L. (2005). Play assessment for group settings: A pilot study to construct an assessment tool. *Scandinavian Journal of Occupational Therapy*, 12, 136–144.
- Lawlor, M.C. (2003). The significance of being occupied: The social construction of childhood occupations. *American Journal of Occupational Therapy*, 57, 424–434.
- Liiva, C.A. & Cleave, P.L. (2005). Roles of initiation and responsiveness in access and participation for children with specific language impairment. *Journal of Speech, Language and Hearing Research*, 48, 868–883.
- Linacre, J.M. (2005) Winsteps, Rasch Model computer program, 3.63.0. Retrieved 30 September 2009, from <http://www.winsteps.com>
- Linacre, J.M. (1987–2008) FASETS, Many-faceted Rasch measurement computer program 1987–2008. Retrieved 4 September 2003, from <http://www.winsteps.com>
- McCabe, P.C. & Marshall, D.J. (2006). Measuring the social competence of preschool children with specific language impairment: Correspondence among informant ratings



- and behavioural observations. *Topics in Early Special Education*, 26, 234–246.
- Miller, C.A. (2001). False belief and sentence complement performance in children with specific language impairment. *Journal of Communication Disorders*, 39, 191–213.
- Nelson, K., Skwerer, K., Goldman, S., Henseler, S., Presler, N. & Walkenfeld, F.F. (2003). Entering a community of minds: An experimental approach to Theory of Mind. *Human Development*, 46, 24–46.
- Pellegrini, A.D., Dupuis, D. & Smith, P.K. (2007). Play in evolution and development. *Developmental review*, 27, 261–276.
- Picone, M. & McCabe, P.C. (2005). The reliability and discriminant validity of the Social Interactive Coding System with language impaired preschoolers. *Journal of Early Childhood and Infant Psychology*, 1, 113–128.
- Rescorla, L. & Goossens, M. (1992). Symbolic play development in toddlers with expressive specific language impairment. *Journal of Speech, Language, and Hearing Research*, 35, 1290–1302.
- Rosnow, R.L. & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, 57, 221–237.
- Tesio, L. (2003). Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *Journal of Rehabilitation Medicine*, 35, 105–115.

### **III**

## **INTER-RATER RELIABILITY OF THE PLAY ASSESSMENT FOR GROUP SETTINGS**

by

Lautamo, T., & Heikkilä, M. (2011)

Scandinavian Journal of Occupational Therapy, 18:1: 3-12

Reproduced with kind permission of the publisher

REVIEW ARTICLE

## Inter-rater reliability of the Play Assessment for Group Settings

TIINA LAUTAMO<sup>1</sup> & MAIJA HEIKKILÄ<sup>2</sup>

<sup>1</sup>JAMK University of Applied Sciences, Rehabilitation and Social Studies, Jyväskylä, Finland, and <sup>2</sup>Helsinki University Hospital, Department of Phoniatics, Helsinki, Finland

### Abstract

Play Assessment for Group Settings (PAGS) was originally designed to measure the play ability of 2- to 8-year-old children in their natural day-care context and for the use of different professionals. The main purpose of this study was to examine the inter-rater reliability of the PAGS. Twelve raters from different professional backgrounds scored a total of 78 children, both videotaped co-scored cases and individual live cases. The many-faceted Rasch model (MFR) was used to examine the goodness-of-fit of the rater severity estimates. Of the 12 raters, 91.7% were demonstrated to fit the MFR model expectations. Overall the results of this study indicated that the raters scored the PAGS in a reliable way, and that they behaved like independent experts. The separation index (2.09) of the raters indicated that the inter-rater scorings were comparable with each other for practical purposes. Further research on the different response processes of children in different diagnostic groups is needed.

**Key words:** *Play ability assessment, test development, inter-rater reliability, many faceted Rasch model, MFR*

### Introduction

Play is the leading occupation of a child and vital for the development of the child's cognitive and social skills (1). Motivation to engage in play is central to the healthy development, occupational balance and well-being of children (2). Play is an occupation in which people engage only because they want to, not because they ought to, in contrast to schoolwork where they might be given strict instructions when to do it and how to do it. Because play can be freely chosen, it may be one of the purest expressions of the human personality (3). It comprises activity with the environment, tools, and interaction with other people. Moreover, enjoyable experiences are essential during play (4). In fact, many social skills and friendships develop through peer interaction (5), and consequently play enables the development of social competence.

Many studies have demonstrated that the play ability of children with disabilities or learning

impairments is significantly weaker than that of their typically developed peers (5–9). Non-typically developed children are also more sensitive to the impact of aspects of the environment (10). For this reason the assessment of children's play ability provides us with very important knowledge about children's performance abilities and the interaction between a child and his/her environment. Assessing children's play abilities should be a fundamental part of professional work in paediatric rehabilitation and education. It is obvious that to assess play suitable instruments are required. We do not obtain enough objective and precise knowledge of children's play ability by observation conducted in the absence of a structured tool (11). The use of valid tools supports the planning of therapeutic or educational interventions and enables their effects to be measured. Thus, the construction of play ability assessment tools (3,11), which are not only valid and reliable but also simple enough to use and interpret when observing play as a many-sided phenomenon, presents a number of challenges.

Correspondence: Tiina Lautamo, JAMK University of Applied Sciences, Rehabilitation and Social Studies, Keskussairaalan tie 21 E, FI-40620 Jyväskylä, Finland.  
E-mail: tiina.lautamo@jamk.fi

(Received 9 October 2008; accepted 10 November 2009)

ISSN 1103-8128 print/ISSN 1651-2014 online © 2011 Informa Healthcare  
DOI: 10.3109/11038120903480048

How to assess and how to have an impact on play is also an important part of occupational therapy practice (12). However, the use of play assessment tools in occupational therapy has been limited because reliable and valid assessment instruments have not been available (13,14). Researchers have developed observation-based play assessment instruments for use in clinical practice such as the Preschool Play Scale, focusing, for example, on participation, tool handling, and imitating (15), and the Child-Initiated Pretend Play Assessment, focusing on symbolic play (7,14). Conversely, it has been argued that the play skills and behaviour of the child should always be observed and assessed in natural play situations (16). Participation in meaningful activity with someone else (17) allows the child's imagination and play ideas in social relations to become observable (8). Nowadays there exist methods of play assessment in a natural environment, such as the Test of Playfulness, which reflects a child's transactions in play content, and the Test of Environmental Supportiveness, which evaluates the impact of environmental aspects on play (6,10,18). The above methods highlight the highly complex, many-sided nature of play and its evaluation. However, despite the complexity of the phenomenon, the method of assessing it should be easy to use by all professionals working daily with children. There was, therefore, a need for a new assessment tool focusing on children's play ability during play in a natural social environment and intended for use by diverse professionals.

The Play Assessment for Group Settings (PAGS) is a new instrument which is grounded in the concepts of theories of occupational therapy (19–23), and the social nature of play (17,24). The theoretical underpinnings of the PAGS combine, on the one hand, the concept of meaningful occupation and a playful attitude, and on the other, the concept of mindful doing and observable organized and imaginative play behaviour. In the PAGS, these components are described as two features: (a) expressing a playful attitude and (b) creating and engaging in play stories. These features are not viewed as separate or hierarchical, but as in dynamic interaction with each other. These theoretical issues are discussed in more detail elsewhere (8,25). The item classification varies from easy to difficult. Each item of the PAGS is scored on a four-point scale indicating the relative amount of time of a child's level of doing: (a) hardly ever, (b) seldom, (c) often, and (d) almost always (8).

By using the PAGS, the play ability of pre-school aged children can be evaluated in their natural social context, for example in day-care. The primary purpose of the PAGS is to observe the play ability of a child in order to establish how a child responds to playful occupational challenges and the possibilities of

social play in his/her natural environment. By using the PAGS, the observer can identify the particular challenges in a child's play performance and implement earlier interventions. The PAGS is meant for use by all professionals (teachers, day-care nurses, and therapists) who work daily with children (8).

The internal scale validity and separation ability of the PAGS have been demonstrated in a previous study (8) with children aged 2–8 years. The person response validity demonstrated less than acceptable goodness-of-fit (8). There are possible reasons for this psychometric problem with the PAGS. One possible reason for the children's misfit scorings is that some of the raters might not have scored the items of the PAGS adequately. The inter-rater reliability of the PAGS has not been studied earlier. However, in the previous study (8) and in an ongoing study (9) we have found that specific play patterns exist among different groups of children, for example, six-year-olds and specific language learning impairments (8,9). In the present study we focus on the inter-rater reliability of the PAGS, and in the study on process (9) we turn our attention to the person response validity of the PAGS in relation to the play patterns of distinct groups of children.

Inter-rater reliability has been the primary target of research on test development in the field of occupational therapy (26), and also one of the most important features of educational and psychological instruments (27). Good inter-rater reliability guarantees that the instrument can be confidently used across raters. Therefore, evaluating inter-rater reliability is an essential step in the development and standardizing of an assessment instrument.

The many-faceted Rasch model (MFR) was developed to take account of the limitations of traditional psychometric statistics. The MFR model is an acceptable and useful method of creating an objective instrument in the psychological sciences. The MFR model provides and enables the construction of objective scales that are separable from the distribution of the attributes of the person it measures (28). Each assessed person ability, item difficulty, and rater severity can be set on the same continuum. These ability measures are comparable with each other, not as sum points, as in traditional reliability methods, but as logit values (27). In addition, the MFR model can take into account errors at the level of the individual person, item, and rater (27,28).

The definitions of validity and reliability in the MFR model differ from the traditional ones (27,28). The specific interest in this study is in evidence based on the response processes of the test items, person ability measures, and rater severity (28). There are advantages in estimating inter-rater reliability by using the MFR model. First, it provides a single statistic that allows for direct comparison of the

severity of all raters (27). We expect raters to be within a reasonable severity range. Specifically, rater severity is independently estimated and transformed into a linear measure of the underlying factor. Rater severity measures are useful for estimating the extent to which systematic differences in severity exist between raters. In addition, while using the MFR model, the differences in rater severity can be adjusted for at an individual and item level in the algorithm that computes participant ability estimates (28). Second, when investigating inter-rater reliability, the raters ought to score the same cases, although the MFR model does not require all the raters to score all the persons or items in order to arrive at an estimate of inter-rater reliability. Therefore, raters may score a particular subset of items or subset of cases, and as long as there is sufficient connectedness across the raters and scorings, it will be possible to directly compare raters (27,29).

Furthermore, the MFR expects raters to behave like independent experts. Too high expectations of agreement in scorings would indicate local dependence among the raters. In fact, the attempt to increase inter-rater reliability can actually reduce the independence of the raters, and so degrade the validity of the measures as measures. In other words, each scoring is expected to provide independent information on the location of the performance of the latent trait. We can use the Rasch Kappa index to calculate the degree of rater dependence (28,30,31).

The main purpose of this study was to examine the cross-professional inter-rater reliability of the PAGS. Specifically, the aim was to examine whether a given rater's scorings of the children's play ability fitted the MFR model for the PAGS (goodness-of-fit) and whether the severity calibrations of the raters are in a reasonable range of severity.

The following research questions were addressed:

- (1) Do the raters score children's play ability consistently enough for the rater severity estimates to demonstrate acceptable goodness-of-fit ( $MnSq$   $0.5 \leq 1.5$ ,  $z < 2$ ) with the MFR model for the PAGS, and are the rater severity estimates associated with reasonable standard errors ( $SE < 0.30$ )?
- (2) Are the calibration differences between the raters within a reasonable range ( $\pm 0.30$ ) and are there cross-professional differences in rater severity?

## Materials and method

### Subjects

The study participants were 12 female raters who used the PAGS to score the play ability of a total of 78 children. The raters of this study had received a professional education in the field of early education (Te), occupational therapy (OT), or physical therapy (PT). They all had earlier experience of observing and supporting children's play (Table I).

The subjects of this study were typically developing children ( $n = 36$ ) and children with learning impairments ( $n = 42$ ) (Table II). The children's ages ranged from 1.8 to 8.8 years ( $M = 5.3$  years). Each of the participants scored six co-scored video cases and six individually selected live cases. The six videotaped co-scored cases were typically developing children. The live cases were both typically developing children and children with learning impairments (see Table I). All the scorings of children's play ability were linked via co-scored video cases.

### Procedure

To allow for co-scoring by other raters, the six children were videotaped during play situations where they were playing freely with familiar peers. The duration of the video case play sessions was 15 to 30 minutes. The six videotaped cases were scored during the two-day training in the use of the PAGS. First, the theoretical background was introduced to the raters and then the principles of scoring the PAGS items. After viewing each video, the challenges of and possible problems in scoring were discussed with the raters in more detail. The raters were not allowed to make any changes in their scorings during the training process.

After the training period and the video case scorings, the participants rated six live cases each in their own workplaces or other locations, i.e. homes. Most of the live cases were assessed in free play situations in social settings, in the day-care centre or at home. Six of the live cases were assessed during therapy sessions while the child was playing with an adult.

Table I. Demographic characteristics of the raters.

Profession	Experience of working with children			Total
	1-5 years	5-15 years	Over 15 years	
Teacher in early education (Te)	1	1	1	3
Occupational therapist (OT)	2	2	3	7
Physiotherapist (PT)	1	1	-	2

Table II. Demographic characteristics of the children by groups.

	Typical ( <i>n</i> = 36) including six video cases (46.2%)	Learning impairments ( <i>n</i> = 42) (53.8%)	Total ( <i>n</i> = 78) (100%)
Gender			
Male	14	24	38 (48.7%)
Female	22	18	40 (51.3%)
Age			
<i>M</i>	5.0	5.5	
Range	2.0–6.9	1.8–8.8	

Consents for all the children in both the videotaped and live cases assessed with the PAGES were provided by their parents. The live case scoring forms were returned without names in envelopes to the researcher.

The study data of this study were collected with the second research version of the PAGES (45 items). The MFR model for the PAGES has been updated according to the principles reported in the previous study (8), and the items had been reduced to 38. The analysis in this study was conducted according to this updated version (3.38) of the PAGES (25).

#### Data analysis

The data comprise three facets (play item difficulty, rater severity, and child ability), and the analysis was implemented by using an MFR model (28). We used the Facets version 3:54.1 (32) program to generate different statistics on how the actual data fitted the expectations of the MFR model for PAGES. In the analysis we used the MFR model for the PAGES based on 38 anchored play items obtained from the MFR analysis of the total database of the play ability scorings of 208 Finnish children, including both typically developed and children with learning problems, aged between 2.1 and 8.9 years of age (*M* = 5.4 years, *SD* = 1.3 years) (33). Further analyses with anchored item values and floating rater severity and child ability were carried out.

First, to determine whether the raters were consistent enough, we examined whether their scorings of children's play ability fitted the MFR model for the PAGES. We expected the data to conform to the following two assertions: (c) the rater is more likely to assign the children lower scores on harder play items and higher scores on easier play items; and (d) lower scores for a child with fewer abilities and higher scores for a child with more abilities. The raters that conform to these expectations will demonstrate

acceptable goodness-of-fit with the MFR model for the PAGES.

The mean square (MnSq) residuals and standardized score (*z*) residuals were used to measure goodness-of-fit of the raters with the model. The criteria for acceptable goodness-of-fit (both infit and outfit) of rater severity were set at MnSq 0.5 to 1.5 and *z* < 2, which is the generally recommended mean square range for productive measurement (34). It is generally expected that 95% of the raters should meet the criterion that 5% may misfit by chance at *z* < 2. Standard error estimates for rater severity measures were investigated. The size of SE is influenced by how well the data fit the model's assertions. The criterion for acceptable standard error for raters was set at SE < 0.30 (35).

We also adopted a second method from a previous study (36) to evaluate inter-rater reliability. We counted the overall proportion of unexpected scorings of raters both in the total sample and separately in the videotaped case scorings. Hence, as the analysis of the total data of 208 Finnish children revealed an overall rate of unexpected scorings of 4%, we set our criterion for an acceptable level of individual unexpected scorings at ≤ 4% when *z* = 2 (33).

To examine whether the raters behaved like independent experts we calculated the Rasch Kappa value which indicates the relation between expected agreement and the observed agreement percentage between the scorings of the raters. In general, we expect the observed percentage agreement to be close to expected percentage agreement. On the whole, the expected value of the Rasch Kappa index is to be positive and close to 0.0. (30).

To answer the second question, i.e. whether the calibration differences between the raters are within a reasonable range, we examined whether the raters maintained the same relative severity. In general, we expected the raters to score items independently, but consistently enough for their scorings to be comparable (30). It is possible for each rater to demonstrate a high level of goodness-of-fit, but be consistently more severe or lenient in relation to all the other raters or of those from a different professional background. According to Linacre (37), calibration stability within ±0.30 logit is the best that can be expected for any variable. However, for practical purposes, Linacre (37) states that variation up to ±0.50 is likely to be free from bias. To confirm the severity of the raters, their calibration values (logits) were expected here to be within ±0.50 logits (37). The separation index is the number of statistically different performance strata that the test can identify in the sample. Therefore, to confirm consistency, the separation index of the raters was expected

to be  $<2.0$  (38). In these analyses the mean of rater severity was centred at zero.

## Results

To answer the question of whether the raters were consistent enough we examined whether they conformed to the expectations of the MFR model with regard to the goodness-of-fit; 11 of the 12 raters (91.7%) demonstrated acceptable goodness-of-fit ( $MnSq >0.5$  or  $<1.5$ ) (Table III). The SE of rater severity was acceptable ( $SE_{mean} = 0.08$ ), which supports the model assertion (Table III). The proportion of individual unexpected scorings was 4.0% when  $z = 2$ , as we expected. In addition, the MFR model analysis revealed that the expected agreement of the rater scorings was 44.1% and the observed agreement of the rater scorings was 50.8%. Consequently, the Rasch Kappa value was 0.12, which was close to the model expectation (0.0). This indicates that the raters in this study were scoring the PAGS items as independent experts (28).

To examine whether the raters maintained the same relative severity we investigated the rater calibration values. In this data the rater calibration values ranged between  $-0.24$  and  $+0.27$  logits, while the mean of rater severity was centred at zero. All the cross-professional raters fitted within 0.51 logits. This was slightly higher than the expected value ( $\pm 0.50$ ) (37). The separation index of the raters was 2.09, which was also slightly higher than the expected  $<2.0$ . The distribution of the raters along a linear scale and targeting the children's play abilities and the PAGS item difficulties is presented in Figure 1.

## Discussion

The main purpose of this study was to examine the cross-professional inter-rater reliability of the PAGS. To do this, we investigated how well the rater scorings data fitted the MFR model for the PAGS, and whether the estimates of the raters were associated with reasonable standard errors. These analyses demonstrated that the raters were scoring the PAGS items in a logical way. Eleven of the 12 raters showed good fit with the assertion of the MFR model. However, one misfitting rater caused a higher percentage of misfits (8.3%) than expected (5.0%). Since this might be a threat to inter-rater reliability, further analysis was undertaken to ascertain whether the scorings of this rater exhibited any systematic patterns. When using the MFR model, it is possible to analyse the scorings of a single rater in detail (28). We found that the most of her misfitting individual scores (73.0%) were among her videotaped case scorings, while her live case scorings were reliable. Therefore, we concluded that this one misfit does not pose a threat to the inter-rater reliability of the PAGS when assessing children's play abilities in natural group settings.

When studying inter-rater reliability of an assessment tool, common cases are required. It has been quite usual to provide for this need by using videotaped situations [see e.g. 2,18,40], as we did here. However, the use of videotapes to observe human occupation has its limitations. For example, the information available to the rater can be restricted and some items might be difficult to identify (40). In the present study, for example, the video operator may have zoomed in on a child playing with toys and thus failed to capture how the children were acting at the

Table III. Rater calibrations (logits) and rater goodness-of-fit ( $MnSq$ ,  $z$ ) to the sMFR model for the PAGS.

Rater		Severity		Infit		Outfit	
		Calibration	SE	$MnSq$	$z$	$MnSq$	$z$
1	PT	0.27	0.08	0.97	-0.3	1.02	0.2
5	OT	0.19	0.08	1.11	1.4	1.21	2.3
8	Te	0.19	0.07	1.01	0.1	1.07	0.9
2	Te	0.18	0.08	0.69	-4.4	0.80	-2.4
4	OT	0.15	0.08	1.09	1.2	1.09	1.1
6	OT	0.04	0.08	1.05	0.6	1.09	0.9
12	Te	0.00	0.10	1.56	6.2	1.58	5.0
11	OT	-0.11	0.07	0.54	-7.9	0.56	-6.8
3	OT	-0.19	0.08	0.59	-6.3	0.61	-5.3
10	OT	-0.23	0.08	0.78	-3.2	0.79	-2.8
7	OT	-0.24	0.08	1.24	3.4	1.27	3.5
9	PT	-0.24	0.10	1.39	4.2	1.42	4.2

Scand J Occup Ther Downloaded from informahealthcare.com by Tina Lautamo on 08/20/12  
For personal use only.

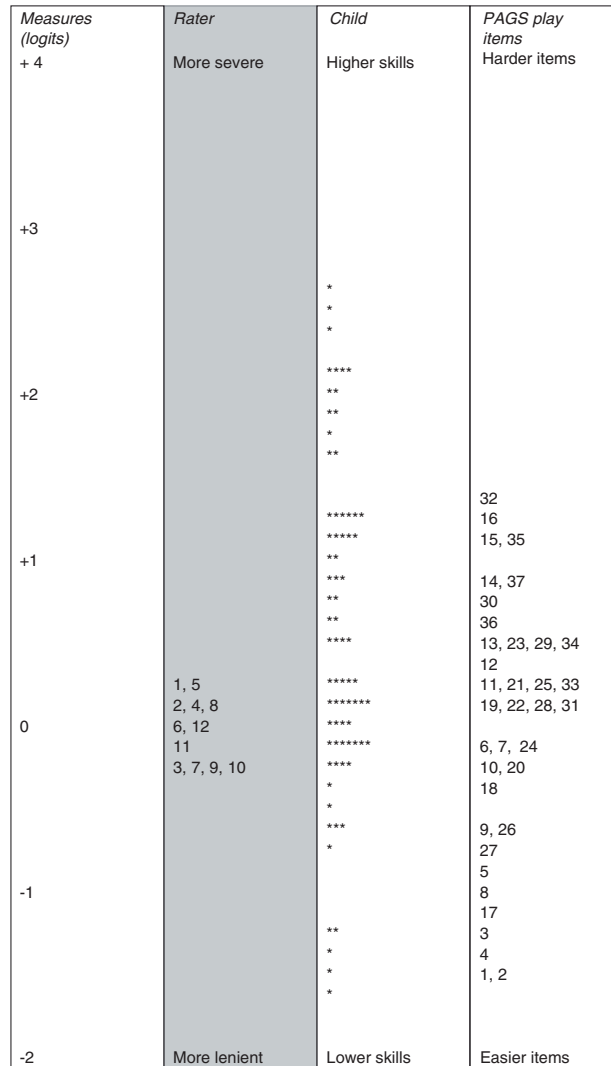


Figure 1. Distribution of raters, play abilities of children, and play items (PAGS version 3.38/2007).

same time. Therefore, in future research it would also be beneficial to organize cases for common assessments in natural social play settings rather than through videos alone.

In this study the raters scored the video cases during the training period and after scoring each item the raters discussed their scorings. The purpose was to intensify their learning. They were not allowed to adjust any scores afterwards, which could have

affected their severity and threatened their ability to behave as independent experts. According to Linacre (31) each scoring is expected to provide independent information on the location of the performance of latent trait, such as play ability. In other words, rater independence is essential for the validity and reliability of the measures.

In this study the rater calibration values ranged around 0.51 logits, which was only slightly higher



than expected. Linacre (37) claimed that for all practical purposes, variation up to  $\pm 0.50$  logits is likely to be free from bias. Linacre (37) also stressed that, as the size of the samples increases, the differences between raters become smaller. We used relatively small sample with six co-scored cases scored on 37 of 38 play items each for a total of 222 ratings, respectively. For a sample of this size, Linacre (37) proposed as a criterion of 99% confidence of calibration stability within  $\pm 0.50$  logits. Furthermore, the separation index of the raters was only slightly higher than expected. The separation index shows that the scorings of the raters separate them no more than into two different levels of severity. These results indicate that the PAGS scorings between the diverse professional raters are probably free from bias and comparable with each other. Thus, the calibration of rater severity is not needed for practical purposes. Consequently, these results support the objective that the PAGS can be used as originally intended by all professionals (teachers, day-care nurses, and therapists) who work on a daily basis with children (8). Nevertheless, when using the MFR model, possible differences in rater severity can be adjusted for on an individual and item level if needed (28). Therefore, rater severity calibration is possible and might be beneficial when using the PAGS to research purposes.

Although the inter-rater reliability of any assessment tool has been an important goal in research only a few studies have used the many-faceted Rasch model to analyse inter-rater reliability. Consequently, the task of creating reliable procedures was particularly challenging. The authors found corresponding methods used in a small number of previous studies [see 2,18,36,41]. However, according to Macmillan (42), the MFR model is comparable to and in fact more accurate than the classical methods needed to study inter-rater reliability. The MFR analysis results in a more accurate evaluation of rater agreement and consistency (27). We believe that using the MFR model in this study of the inter-rater reliability of the PAGS gave us more information than would have been obtained using other, more traditional, methods.

### Conclusion

The results of this study indicate that raters score the PAGS in a reliable way. The results of the present and previous study support both the scale validity and inter-rater reliability of the PAGS as an objective method for evaluating play in natural settings. However, the challenge remains to investigate the person response validity bias of the PAGS. We need to confirm whether specific groups of children exhibit specific play patterns by comparing the profiles of

individual play items. This will be the topic of our next article and will be based on a large database.

### Acknowledgements

The authors would like to warmly thank all the raters, children, and their parents who committed to this study. They would also like to thank Michael Freeman for checking the language of this manuscript. This study was supported by grants from the Finnish Association of Occupational Therapists.

### References

1. Parham LD, Primeau LA. Play and occupational therapy. In: Parham LD, Fazio LS, editors. *Play in occupational therapy for children*. St Louis, MO: Mosby; 1997. p. 3–21.
2. Andersen S, Kielhofner G, Lai J-S. An examination of the measurement properties of the Pediatric Volitional Questionnaire. *Phys Occup Ther Ped* 2005;25:39–57.
3. Bundy AC. Assessment of play and leisure: Delineation of the problem. *Am J Occup Ther* 1993;47;3:217–22.
4. Hinojosa J, Kramer P. Integrating children with disabilities into family play. In: Parham LD, Fazio LA, editors. *Play in occupational therapy for children*. St Louis, MO: Mosby; 1997.
5. Leipold EE, Bundy AC. Playfulness in children with attention deficit hyperactivity disorder. *Occup Ther J Res* 2000;20:61–82.
6. Okimoto AM, Bundy A, Hanzlik J. Playfulness in children with and without disability: Measurement and intervention. *Am J Occup Ther* 2000;54;1:73–82.
7. Stagnitti K, Unsworth C, Rodger S. Development of an assessment to identify play behaviours that discriminate between the play of typical preschoolers and preschoolers with pre-academic problems. *Can J Occup Ther* 2000;67: 291–303.
8. Lautamo T, Kottorp A, Salminen A-L. Play assessment for group settings: A pilot study to construct an assessment tool. *Scand J Occup Ther* 2005;12:136–44.
9. Lautamo T, Laakso M-L, Aro T, Ahonen T. Play patterns in children with specific language impairment assessed in natural group settings. *Forthcoming* 2010.
10. Rogers Bronson M, Bundy AC. A correlational study of a test of playfulness and a test of environmental supportiveness for play. *Occup Ther J Res* 2001;21:241–59.
11. Sturgess JL. Current trends in assessing children's play. *Br J Occup Ther* 1997;60:410–14.
12. Couch KJ, Deitz JC, Kanny EM. The role of play in pediatric occupational therapy. *Am J Occup Ther* 1998;52: 111–17.
13. Stagnitti K. Understanding play: The implications for play assessment. *Aust J Occup Ther* 2004;51:3–12.
14. Stagnitti K, Unsworth C. The test-retest reliability of the child-initiated Pretend Play Assessment. *Am J Occup Ther* 2004;58:93–9.
15. Knox S. Development and current use of the Knox Preschool Play Scale. In: Parham LD, Fazio LA, editors. *Play in occupational therapy for children*. St Louis, MO: Mosby; 1997, p. 35–51.
16. Bundy AC. Measuring play performance. In: Law M, Baum C, Dunn W, editors. *Measuring occupational performance*. Thorofare, NJ: Slack; 2001. p. 89–102.

17. Lawlor MC. The significance of being occupied: The social construction of childhood occupations. *Am J Occup Ther* 2003;57:424–34.
18. Bundy AC, Nelson L, Metzger M, Bingaman K. Validity and reliability of a test of playfulness. *Occup Ther J Res* 2001;21:276–92.
19. Bundy AC. Play and playfulness: What to look for. In: Parham LD, Fazio LS, editors. *Play in occupational therapy for children*. St Louis, MO: Mosby; 1997. p. 52–66.
20. Ferland F. *PLAY, children with physical disabilities and occupational therapy: the Ludic Model*. Ottawa: University of Ottawa Press; 1997.
21. Christiansen CH. Defining lives: Occupational identity: An essay on competence, coherence, and the creation of meaning. *Am J Occup Ther* 1999;53:547–58.
22. Bundy AC, Murray EA. Sensory integration: A. Jean Ayres theory revised. In: Bundy AC, Lane SJ, Murray EA, editors. *Sensory integration, theory and practice*. 2nd ed. Philadelphia: FA Davis; 2002. p. 227–40.
23. Kielhofner G. *Model of human occupation: Theory and application*. 3rd ed. Baltimore, MD: Williams & Wilkins; 2002.
24. Leslie AM. Pretense and representation: The origins of “Theory of Mind”. *Psychol Rev* 1987;94:412–26.
25. Lautamo T. Assessing play in a social setting. In: Stagnitti K, Cooper R, editors. *Play as therapy: Assessment and therapeutic interventions*. London: Jessica Kingsley; 2009. p. 115–29.
26. Ottenbacher KJ, Tomchek SD. Reliability analysis in therapeutic research: Practice and procedures. *Am J Occup Ther* 1993;47:1–16.
27. Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability: Practical assessment, research and evaluation. *PAREonline.net* 2004 [cited 20 July 2009]. Available at: <http://www.pareonline.net/getvn.asp?v=9&n=4>
28. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental measurement in human sciences*. Hillsdale, NJ: Lawrence Erlbaum; 2007.
29. Linacre JM, Englehard G, Tatem DS, & Myford CM. Measurement with judges: Many-faceted conjoint measurement. *Int J Educ Res* 1994;21:569–77.
30. Linacre JM. A user’s guide to FASETS. [cited 16 March 2007]. Available at: <http://winsteps.com/winman/reliability.htm>.
31. Linacre JM. Judge ratings with forced agreement. *Rasch Measurement Trans* 2002;16:857–8.
32. Linacre JM. Many-faceted Rasch measurement computer program 1987–2004. Available at: <http://www.winsteps.com>.
33. Lautamo T. The play ability measures for normally developing children and children with learning impairments between 2 and 8 years of age. Unpublished database; 2009.
34. Wright B, Linacre M. Reasonable mean-square fit values. *Rasch Measurement Trans* 1994;8:3 [cited 30 June 2009]. Available at: <http://www.rasch.org/rmt/rmt83b.htm>
35. Tham K, Bernspång B, Fisher AG. Development of the Assessment of Awareness of Disability. *Scand J Occup Ther* 1999;6:184–90.
36. Coto S, Fisher A, Mayberry WL. The assessment of motor and process skills applied cross-culturally to Japanese. *Am Occup Ther* 1995;21:798–806.
37. Linacre JM. Sample size and item calibration (or person measure) stability. *Rasch Measurement Trans* 1994;7:328 [cited 12 July 2009]. Available at: <http://www.rasch.org/rmt/rmt74m.htm>
38. Wright BD. Reliability and separation. *Rasch Measurement Trans* 1996;9:472 [cited 16 March 2009]. Available at: <http://www.rasch.org/rmt/rmt94n.htm>
39. Shutton-Smith B. *The ambiguity of play*. Cambridge, MA: Harvard University Press; 1997.
40. Hermansson LM, Bodin L, Eliasson A-C. Intra- and inter-rater reliability of the assessment of capacity for myoelectric control. *J Rehabil Med* 2006;38:118–23.
41. Atchison BT, Fisher AG, Bryze K. Rater reliability and internal scale and person validity of the School Assessment of Motor and Process Skills. *Am J Occup Ther* 1998;52: 843–50.
42. Mcmillan P. Classical, generalizability, and multifaceted Rasch detection of interrater reliability in large, sparse data sets. *J Exp Ed* 2000;68:2. Available at: <http://web.ebscohost.com>

## IV

### **TEST-RETEST RELIABILITY OF THE PLAY ASSESSMENT FOR GROUP SETTINGS (PAGS)**

Corresponding author:

Tiina Lautamo, OT Reg, LicHSc, PhD student

Department of Rehabilitation and Social Sciences, JAMK University of Applied  
Sciences and Department of Educational Science, University of Jyväskylä

Marja-Leena Laakso, PhD

Department of Educational Science, University of Jyväskylä

Tuija Aro, PhD

Department of Psychology, University of Jyväskylä and Niilo Mäki Institute

Timo Ahonen, PhD

Department of Psychology, University of Jyväskylä

Manuscript submitted in American Journal of Occupational Therapy  
in June 2012

## **ABSTRACT**

**Aim:** The main focus was to evaluate the test-retest reliability of the PAGES.

**Methods:** Six occupational therapy students scored 29 typically developing children's aged 3 to 5.5 years play performance twice at a systematic one-week interval. The many-faceted Rasch model (MFR) was used to examine the consistency and severity estimates of the raters. The stability of the individual measures was investigated with the intra-class correlation (ICC) coefficient.

**Results:** Five of the six raters scored the PAGES in a logical way. The rater separation index was reasonable but the number of unexpected scorings was slightly higher than expected. The ICC for the single measures of the children was weak but for average measures good.

**Conclusion:** Some uncertainty among the inexperienced raters together with small sample size might have caused the fairly weak ICC for the single measures. We therefore need to reevaluate the intra-rater reliability with a larger sample.

**Keywords:** Play assessment, test-retest reliability, rater consistency, rater severity

## INTRODUCTION

Play is seen as a primary and important activity of children and the way to growth and development towards agency and participation (Lawlor, 2003). Therefore, play has been of interest to many authors for several decades (e.g. Bundy, 1993; Eisert & Lamorey, 1996; Hartley, 1963; Kooij, 1989; Lieberman, 1977; Parham, 2008; Shutton-Smith, 1997). Much of children's early learning is achieved through play. Their physical and socio-emotional and intellectual development is dependent on their ability to play with others (Parham, 2008). Difficulties in play performance in social environments like daycare may give rise to further problems in learning and participation (Pellegrini, Dupuis, & Smith, 2007). Thus, the assessment of children's play performance provides us with important information on children's performance abilities and should therefore be a fundamental component of professional working with children in early education and rehabilitation (Kelly-Vance & Ryals, 2007).

The Play Assessment for Group Settings (PAGS) has been developed for use by daycare and rehabilitation professionals in evaluating children's play performance in group settings. The PAGS is originally meant to measure the play performance of two- to eight-year-old children (Lautamo, Kottorp, & Salminen, 2005; Lautamo & Heikkilä, 2010; Lautamo, Laakso, Aro, Ahonen, & Törmäkangas, 2011). The PAGS is an occupation science-based assessment instrument. It is intended to be used as a structured tool to observe the play performance of a child so as to evaluate his/her responses to playful occupational challenges and social play opportunities in daily play settings (Lautamo, 2009). The practical usefulness of the PAGS concerns not only whether we can use the results of the observations to identify children in need of support, such as individual intervention or consultation services, in their play performance, but also whether the PAGS can be used to evaluate changes and document the effectiveness of possible interventions. Moreover, if we wish to measure post intervention changes in a child's play performance, we need a tool with good test-retest reliability, i.e. with short-term temporal stability. Therefore, in this study we focus on examining the test-retest reliability of the PAGS.

In previous studies, the many-faceted Rasch model (MFR) has been used to yield evidence on the validity of the internal structure of the PAGS, the relation of the PAGS to other variables, and child response and rater response processes. The many-faceted Rasch model (MFR) was developed to take account of the limitations of traditional psychometric statistics. The MFR model enables the construction of objective scales that are separable from the distribution of the attributes of the person it measures (Bond & Fox, 2007). Moreover, the MFR model can take into account errors at the level of the individual person, item, and rater (Bond & Fox, 2007; Stemler, 2004). When estimating a scored phenomenon, in this case play performance, with the MFR model, all the test items are expected to form a unidimensional continuum

representing the theoretical construct of the phenomenon being measured (Bond & Fox, 2007; Tesio, 2003). The results of earlier studies (Lautamo et al., 2005, 2011) have supported the internal structure (internal scale validity) of the PAGES. In other words all the scale items were well distributed along a linear scale and targeted to the play ability of the children, and all worked together to represent the children's play performance. Moreover, it is critical that a valid assessment of play performance differentiates between persons with limitations and those without limitations. Earlier studies (Lautamo et al., 2005, 2011) have confirmed the validity of the PAGES in relation to other variables, i.e., the PAGES separated the children with low play performance ability from those with high play performance ability.

When using MFR analysis it is generally expected that the test items will remain stable, but shift upward or downward in proportion to the overall mean difference in values across the different groups being studied. However, when we investigated the differentially functioning items (DIF) across two groups of children, i.e. children with and without specific language impairment (SLI), we found some inconsistency in the PAGES scale. In fact, our previous study (Lautamo, et al., 2011) showed seven items out of 38 to differentiate in relative challenge. Five of these seven items proved relatively harder for the children with SLI, mainly because they demanded the use of language skills in the play situation.

Furthermore, slightly too much variance was found in person response processes in the play performance scorings of the PAGES (Lautamo et al., 2005). Data from approximately 10% of the children failed to conform to the expectations (5 %) of the MFR model for the PAGES (Lautamo et al., 2005). This means that the play performance of more children than expected did not conform to the following two assumptions: (a) the child is more likely to obtain higher scores on easier play items than on harder play items; and (b) the easier play items are more likely to be easier for all children than the harder play items. A possible reason for this person response variance might be the previously revealed differential item functioning (DIF) (see Decker, Dallmeijer & Lankhorst, 2005); more research is clearly needed to establish the person response process validity of the PAGES in different groups of children, e.g. cultural or diagnostic groups. However, despite the slight uniformity and excessive variance in the children's response processes, we found (Lautamo, et al. 2011) that the PAGES was nevertheless able to separate children with observable or diagnosed developmental challenges from those typically developing in their play performance (Lautamo et al., 2005, 2011). While the children's response processes can vary, this may also cause estimation bias when evaluating other aspects of the validity and reliability of the tool. Therefore, in order to minimize the variability in individual children's play performance, we chose a fairly homogenous group of children; typically developing children within a limited age range, i.e. from 3 years to 5 years 6 months, to investigate test-retest reliability in the present study.

According to Schumacker (2007) a previously reported estimate of internal consistency of a test is not informative unless the proposed sample has exactly the same score distribution as the sample used for the reported internal consistency. The MFR model offers an anchoring technique to attain item hierarchy consistency (Bond & Fox, 2007). Therefore, in investigating rater reliability and test-retest reliability with a fairly small sample of children, we used the anchored item values (taken from the larger PAGES database) to determine whether the raters fit the MFR model expectations. Moreover, instead of relying on previously reported estimates of inter-rater reliability (rater response processes) (Lautamo & Heikkilä, 2010) we confirmed rater consistency and severity for the present data before estimating test-retest reliability. We used the MFR model for this purpose, since it provides more detailed output on the basis of rater severity and rater consistency and thus enables better understanding of judge rating behavior and judge bias in ratings (Bond & Fox, 2007).

The MFR model provides a single statistic that allows for direct comparison of the severity of all raters (Stemler, 2004). Specifically, rater severity is independently estimated and transformed into a linear measure. We expected the raters to be within a reasonable severity range so that their ratings would be comparable to each other. The MFR model does not require all the raters to score all the persons or items in order to arrive at an estimate of inter-rater reliability. Therefore, raters may score a particular subset of items or subset of cases, and as long as there is sufficient connectedness across the raters and scorings, it will be possible to directly compare raters (Linacre, et al., 1994; Stemler, 2004). Furthermore, the raters are expected to behave like independent experts. Too high expectations of agreement in scorings would indicate local dependence among the raters. In fact, attempts to increase inter-rater reliability can actually reduce the independence of the raters, and so degrade the validity of the measures. In other words, each scoring is expected to provide independent information on the location of the performance of the latent trait (Bond & Fox, 2007). In the MFR model each assessed person ability, item difficulty, and rater severity can be set on the same continuum. All the measures are comparable with each other as logit values (Stemler, 2004). Since the MFR analyses yield comparable total play performance logit values for each child's level of play performance in the form of additive numbers (logits), we are able to investigate test-retest reliability with more traditional statistical methods. In the fields of education and psychology, the term reliability is operationalized as relative consistency and quantified using reliability coefficients called intraclass correlation coefficients (ICCs) (Weir, 2005)

Test-retest reliability refers a measure of consistency of a test across time. The length of time between the tests is crucial as it affects test-retest reliability. A very short time interval makes carryover effects due to memory, practice, or mood more likely, whereas a longer interval increases the chances that a change in status could occur (Allen & Yin, 1979). Nevertheless, the possibility exists that a change in scores represents a real change in the true score rather than

random error. However, test-retest reliability can also capture sources of error external to the actual instrument. Longer periods in particular matter because people are more likely to change over the time interval, while a shorter time can lead to learning effects (Allen & Yin, 1979). When assessing play performance as a daily occupation in a natural learning environment, there is no threat of other than natural learning effects and, therefore, we decided to choose a short (one-week) interval for testing.

The main purpose of this study was to examine the test-retest reliability of the PAGES by comparing single measures of the children's play performance conducted at a one-week time interval. We wished to confirm, first, the inter-rater reliability in the PAGES data by examining rater goodness of fit to the MFR model and rater severity when the raters were scoring children's play performance in social non-structured play situations. The specific research questions were:

1. Do the raters score children's play performance consistently enough for the rater severity estimates to demonstrate acceptable goodness of fit ( $MnSq \geq 0.5$  and  $\leq 1.5$ ,  $z < 2$ ) with the MFR model for the PAGES, and are the rater severity estimates associated with reasonable standard errors ( $SE < 0.30$ )?
2. Are the calibration differences between the raters within a reasonable range ( $\pm 0.50$ )?
3. Do individual measures of the children's play performance remain stable when measured twice at a one-week interval ( $ICC > 0.60$ )?



## MATERIALS AND METHODS

### Participants

The sample in this study consisted of the play performance ratings of six raters and twenty-nine children. The raters were female second-year occupational therapy students with only little previous experience in observing of children's play performance. The children ( $n=29$ ) were typically developing boys ( $n=12$ ) and girls ( $n=17$ ) aged 3 years to 5 years 6 months ( $M= 4.5$  years) and they were rated twice at a one-week interval.

### Procedures

Prior to commencement of this study ethical approval was obtained from the manager of the daycare services in the town where the children attended day care. Consent for children's participation was given by their parents. For the present data collection the children's play performance was observed in two different day care centers in free play situations, where the child had an opportunity to join in play with other children.

All the observations and ratings were completed by second-year occupational therapy students. The raters were provided with two one-day training sessions, including theoretical reflection and practical training. The raters practiced by doing pre observations of two children in live play situations. After the training assessments, the raters were tutored in reflective group discussions. Pair of raters rated the same 10 children (blinded to each other). The follow-up ratings were done at a one-week interval following the same procedure. In order to connect all the data together for the MFR analyses the ratings were done in pairs (T1) and for the second rating (T2) one rater of the pair was changed (Table 1).

Insert TABLE 1 here

Ratings were done with the Play Assessment for Group Settings (PAGS) current version (version 3) which consists of 38 play items (Lautamo, 2009). When evaluating play performance, the raters used time-related scoring, describing the approximate amount of time the child's play behaviour reflected a certain play item during a free play situation. Each of the 38 items of the PAGS was scored on a 4-point scale indicating the relative amount of time that the child's play behaviour reflected that item. The time-related scores are total approximations of the child's play performance in different play situations. The raters were encouraged to make observations in at least two different play situations. An observation assessment session lasted approximately 45 minutes per child. If the child needed time to initiate the play performance, the raters were allowed to continue the observation time to acquire enough information for the ratings.

Day care centers in Finland are environments especially planned for child day care purposes. Children commonly spend approximately 8 hours a day in day care, and have several opportunities for free play with various toys. Day care group size can vary, but in the observed play situations small groups of children played together (2 to 5 children) or some children played alone by their own choice. All observations were made when the children were playing inside the center building.

### **Data analyses**

In order to generate statistics on how the actual data fitted the expectations of the MFR model for the PAGS, we used the previously modeled and anchored 38 play items (Lautamo, 2009). Anchoring refers to a procedure that specifies that the present values are to be used in the analysis. The anchoring values were obtained from the MFR analysis of the total database of the play performance scorings of 208 Finnish children, including both typically developing and children with learning impairments, aged between 2.1 and 8.9 years of age ( $M = 5.4$  years,  $SD = 1.3$  years) (unpublished database). In these analyses we used FACETS, version 3.54.1, a MFR analysis computer program (Linacre, 1987-2008).

The data for the present study comprised three facets (play item difficulty, rater severity, and child ability). The PAGS rating form allows the rater to leave blank items which refer to play behavior not observed in the rating situation. In the present data, three items (item 10, 31 and 32) had only a few ratings. These three items were removed from the further analyses. Further analyses with anchored 35 item values and floating rater severity and child ability were conducted.

First, to determine whether the raters were consistent enough, we examined whether their scorings of the children's play performance fitted the MFR model for the PAGS. We expected the data to conform to the following two assumptions: (a) the rater is more likely to assign the children lower scores on harder play items and higher scores on easier play items; and (b) to assign a child with fewer abilities lower scores and a child with more abilities higher scores. Raters who conform to these expectations will demonstrate acceptable goodness of fit with the MFR model for the PAGS. To measure goodness of fit of the raters the mean square (MnSq) residuals and standardized score ( $z$ ) residuals were used. The criteria for acceptable goodness of fit (both infit and outfit) of rater severity were set at  $MnSq \leq 0.5$  and  $\leq 1.5$ , and  $z < 2$ , which is the generally recommended mean square range for productive measurement (Wright & Linacre, 1994). It is generally expected that 95% of the raters should meet the criterion. Standard error estimates for the rater severity measures were calculated. The size of the SE is influenced by how well the data fit the model's assertions. The criterion for an acceptable standard error for the raters was set at  $SE < 0.30$  (Tham, Bernspång, & Fisher, 1999).

Secondly, we investigated whether the calibration differences between the raters were within a reasonable range and whether the raters maintained the same relative severity. In general, we expected the raters to score items

independently, but consistently enough for their scorings to be comparable. It is possible for each rater to demonstrate a high level of goodness of fit, but be consistently more severe or lenient than the other raters. According to Linacre (1994), calibration stability within  $\pm 0.30$  logits is the best that can be expected for any variable. However, for practical purposes Linacre (1994) states that variation up to  $\pm 0.50$  is likely to be free from bias. To confirm the severity of the raters, their calibration values (logits) were expected here to be within  $\pm 0.50$  logits. The separation index is the number of statistically different performance strata that the test can identify in the sample. Therefore, to confirm consistency, the separation index of the raters was expected to be  $< 2.0$  (see Wright & Linacre 1994). In these analyses the mean of rater severity was centered at zero. In these analyses one rater (rater 2) demonstrated too much leniency and her ratings of the children's play performance were removed from the data.

To examine whether the five reminding raters behaved like independent experts, we calculated the Rasch Kappa value which indicates the relation between expected agreement and the observed percentage agreement between the scorings of the raters. In general, we expect the observed percentage agreement to be close to the expected percentage agreement. On the whole, the value of the Rasch Kappa index is expected to be positive and close to 0.0. (Linacre, 2011). We also adopted a second method from a previous study (Coto, Fisher, & Mayberry, 1995) to evaluate inter-rater reliability further. We counted the overall proportion of unexpected scorings of the raters both in the total sample and separately in the videotaped case scorings. Hence, as the analysis of the total data on 208 Finnish children revealed an overall rate of unexpected scorings of 4%, we set our criterion for an acceptable level of individual unexpected scorings at  $< 4\%$  when  $z = 2$  (Tham, et al., 1999).

Finally, we converted the ordinal raw scores (scored by five raters) of the individual children's ( $n=29$ ) play performance to play performance measures (logits) for two different observation occasions at a one-week interval. To examine the test-retest reliability of the PAGS, the children's two individual (T1 and T2) play performance measures (logits) were inserted into the SPSS18.0 program for further analysis. Since the time interval between two scorings of the individual child was exactly one week, there is only one factor that can be expected to represent a systematic source of variability, i.e. time, and hence one-way ANOVA was used (see Schuk, 2004) to assess the interclass correlation coefficient (ICC). ICCs can vary from 0.000 to 1.00 but according to Weir (2005) there is no consensus on what constitutes a good ICC. However, the commonly used values of from 0.60 to 0.80 are regarded as evidence of good reliability, those above 0.80 indicating excellent reliability. Consequently, we need not only quantitative but also qualitative evaluation of the magnitude of the ICC (Weir, 2005).

## RESULTS

The second MFR analysis was conducted to determine rater consistency. The five reasonable raters' scorings fitted the expectations of the MFR model ( $MnSq \geq 0.5$  or  $\leq 1.5$ ) (Table 2). The SE of the rater severity measures was acceptable ( $SE_{mean} = 0.06$ ), which supports the model assumption. In addition, the MFR analysis revealed that the expected agreement of the rater scorings was 38.1% and the observed agreement of the rater scorings was 49.8%. Consequently, the Rasch Kappa value was 0.17, which was relatively close to the model expectation (0.0). This indicates that the five raters in this study scored the PAGES items not only consistently but independently (Bond & Fox, 2007). However, the proportion of individual unexpected scorings was 5.6 % ( $z = 2$ ), which was slightly higher than we expected ( $< 4.0$  %). No systematic error pattern for specific raters was found; instead, three items (item 7, 16.9 %, item 19, 11.6 %, and item 22, 19.8 %) explained 48.3 % of the unexpected responses. These unexpected scorings are discussed below.

Further investigation of whether the calibration differences between the five raters were within a reasonable range showed that the rater calibration values ranged between -0.24 and + 0.29, while the mean of rater severity was centered at zero. All five raters fitted within  $\pm 0.53$  logits. This was close to our expectation ( $\pm 0.50$ ) (see Linacre, 2011) but the separation index for the raters was 2.93, which was higher than expected  $< 2.0$ . The rater calibrations (logits) values and rater goodness of fit ( $MnSq$ ,  $z$ ) statistics are presented in Table 2. We decided to retain all five raters' ratings in the further analyses. By so doing, we confirmed that all these ratings were connected to the MFR model and that all the children's performance measures were comparable to each other.

Insert Table 2 here

Investigation of the test-retest reliability of the PAGES at a one-week interval revealed an ICC of 0.481, (95 %,  $CI$  0.15 - 0.72,  $p < 0.01$ ) for single measures, and of 0.649 (95 %,  $CI$  0.26 - 0.83,  $p < 0.01$ ) for average measures. The ICC values indicate that the test-retest reliability was good on the group level but fairly weak on the individual level. In fact, 14 of the 29 children's play performance measures differed by more than 0.50 logit and six by more than 1.0 logit (Table 3).

Insert Table 3 here

## DISCUSSION

The main purpose of this study was to investigate the test-retest reliability of the PAGS. To determine whether the children's play performance measures remained stable when repeat-measured at a one-week interval, we calibrated their play performance responses. In contrast to other reliability estimates, the test-retest reliability or reproducibility of a test captures not only the measurement error of an instrument, but also the stability of the construct being measured (Schuck, 2004).

First, we determined whether the raters were consistent enough and their ratings within a reasonable range of severity. To do this, we investigated how well the raters' scorings fitted the MFR model for the PAGS, and whether the rater estimates were associated with reasonable standard errors. These analyses demonstrated that five of the six raters scored the PAGS items in a logical way. One of the raters was far more lenient than the others; after removing this lenient rater's scorings, the remaining raters showed even better fit with the assumption of the MRF model, although they continued to vary slightly in severity. While the proportion of individual unexpected scorings remained higher than expected, we found no systematic error concerning any specific rater. The slight uncertainty in the rating behavior of the student raters compared to that of the professionals (in early education and rehabilitation) (Lautamo et al., 2005; 2011; Lautamo & Heikkilä, 2010), despite training of the former group in the data collection, might be caused by the student raters limited prior experience in observing play performance.

Further investigation of the proportion of unexpected scorings revealed that three items (item 7, item 19, and item 22) explained 48.3 % of the unexpected responses. Since these items were well functioning in the total database of the MFR model for the PAGS, we concluded that the reason was specific to the present data collection. Children were scored down in these three items as compared to their other play performance scores and total play performance level. The possible reasons for low scores were either that the play behavior in question did not occur sufficiently during the observation occasions or that the raters did not have clear enough instructions on how to observe this kind of play behavior. Low scores on item 7 (plays a trick and makes jokes) might reflect the possibility that the presence of an unfamiliar observer affected the play situation or that the overall orientation towards playfulness was weak. Low scores on item 19 (attributes qualities to objects, e.g. the doll is ill, the car is broken) and item 22 (replaces missing objects with imaginary ones, e.g. pretends there is food in the pot) reflect symbolic actions by the child in play. It might be that the fairly short observation period did not include object substitution or attribution of object qualities, despite the child's ability to make symbolic use of objects. The toy repertoire available to the child may also affect the amount of symbolic play; however, this was not controlled for in our study.

Evaluation of the test-retest responses at a one-week interval revealed that the ICC for single measures (0.481) can be considered to be weak, while the

ICC for average measures was better (0.649). Thus there was too much variance between the two different play performance measures in this study. These findings pose a potential threat to the repeated measures reliability of the PAGES. In fact, half of the individual children's play performance measures varied by more than 0.50 logit while six of them varied by more than 1.0 logit across the one-week interval. According to Linacre (1994), calibration stability within  $\pm .30$  logit is the best that can be expected for any variable, and that, for all practical purposes, variations up to  $\pm .50$  logit are likely to be free from bias. However, Linacre (1994) also stressed that calibration estimation error is dependent on the size of the sample used to calibrate the items, persons, or raters. Owing to time constraints, we used a relatively small sample: 29 cases scored on 35 items. As a criterion for sample sizes such as this Linacre (1994) proposed 95% confidence of calibration stability within  $\pm 1.0$  logit. While Linacre (1994) argued that the 30 cases for 30 items might be enough for stable measures, he recommended use of an empirical dataset to the center of the distribution of the 100 to obtain more credible findings. In accordance with these arguments we can interpret our small sample to be stable within  $\pm 1.0$  logit. This indicates that in larger samples the ICC would not necessarily be poor. Therefore, there is a need to reevaluate the test-retest reliability of the PAGES with a larger sample and controlling for rater severity to be able to determine if the PAGES can be used as a reliable tool to measure changes in children's play performance.

It is widely known that the play is flexible and fragile phenomenon and easily disturbed (e.g. Bundy, 1997; Collier & Bundy, 2009). Cordier and Bundy (2009) depicted playfulness as a spider's web, not only to describe its fragility but also its efficacy in capturing players. As a construct, play performance can be seen as changeable rather than stable, owing to the strong effect of both social and physical environmental factors. According to the present raters' comments (notes on the scoring forms), a child's play performance was dependent on environmental factors; for example, if a child's best playmate was absent because of illness, the child was unwilling to initiate play without the intervention of an adult. Likewise, mood affected to the attention a child directed towards the environment; for example, during the first observation a tired child preferred to play alone, and concentrate on construction play, while during the second observation the same child was very eager to join a more complex social play scenario. These changes in individual children's play might have caused the higher variations in the play performance measures separated by a one-week interval. Therefore, it is important that when they observe children, raters confirm that the level of play performance is optimal by choosing an observation time when the child is lively and the environment favorable for free play.

In our study the pair of inexperienced occupational therapy students with limited experience in observing children's play performance evaluated the children on the basis of a 45-minute observation occasion that they had fixed beforehand. This means that neither the selected time nor environmental aspects were necessarily optimal for assessing the child's potential play

performance. One week later, the raters re-evaluated the same child but with a different partner. Since only one of the pair of raters observed the same child it is possible that the slight variability in rater severity in this study affected the child's evaluation from one time to the other. Our findings indicate that when evaluating play performance, a better result is more likely if the rater knows the child and his/her play style and if the observations are done in optimal play situations. Lengthening the observation time is not necessary; instead a diversity of play possibilities and familiarity with the environment are crucial. Brentnall et al. (2008), in scoring children's playfulness with the Test of Playfulness, an observation-based rating scale tool, found that a longer observation time (30 min) , did not necessarily provide a more representative picture than a shorter time (15 min). However, it has been argued that the play ability of a child should always be observed and measured in natural settings (Bundy, 2001) which are safe and familiar enough to support play performance and enable a playful attitude (Pellegrini et al., 2007). Environments that challenge a child's capacities tend to evoke involvement, attentiveness, and maximal activity (Csikszentmihalyi, 1990), but it must be recognized that natural peer settings are not necessarily equally challenging environments for all children and that the nature and level of challenge can vary from day to day.

We concluded that the slight difference in rater severity and some uncertainty among the trained but inexperienced raters together with a small sample size might have caused the fairly weak ICC in the PAGES play performance measures. We need, therefore, to reevaluate the intra rater reliability of the PAGES with a larger sample and experienced raters.

## REFERENCES

- Allen, M.J., & Yen, W.M. 1979. Introduction to measurement theory. Monterey (CA): Brooks Cole.
- Bond, T.G & Fox, C.M. 2007. Applying the Rasch Model: Fundamental measurement in human sciences. Hillsdale, NJ: Lawrence Erlbaum.
- Bundy, A.C. 1993. Assessment of play and leisure: Delineation of the problem. *American Journal of Occupational Therapy*, 47: 217-22.
- Bundy, A.C. 1997. Play and playfulness: What to look for. In L.D. Parham, & L.S. Fazio (eds.) Play in occupational therapy for children (pp.52-66). St. Louis: Mosby.
- Bundy, A.C. 2001. Measuring play performance. In M. Law, C. Baum, & W. Dunn (eds.) Measuring occupational performance (pp.89-102). Thorofare: SLACK
- Csikszentmihalyi, M. 1990. Flow: The psychology of optimal experience. New York: Harper & Row
- Collier, R. & Bundy, A. 2009. Children and Playfulness In: K. Stagnitti & R. Cooper (Eds.), Play as therapy: Assessment and therapeutic interventions (pp. 45-58). London: Jessica Kingsley.
- Coto, S., Fisher, A., & Mayberry, W.L. 1995. The assessment of motor and process skills applied cross-culturally to Japanese. *American Journal of Occupational Therapy*, 21, 798-806.
- Decker, J., Dallmeijer, A.J. & Lankhorst, G.J. 2005. Clinimetrics in rehabilitation medicine: Current issues in developing and applying measurement instruments. *Journal of Rehabilitation Medicine*, 37, 193-201.
- Eisert, D., & Lamorey, S. 1996. Play as a window on child development: The relationship between play and other developmental domains. *Early Education and Development*, 7, 221-235.
- Hartley, R.E. & Goldenson R.M. 1963. The complete book of children's play. New York: Crowell.
- Kelly-Vance, L. & Ryalls, B.O. 2007. Best Practices in Play Assessment and Intervention. In A. Thomas & J. Grimes (Eds), Best Practices in School Psychology V. (pp. 549-559). Bethesda: National Association of School Psychologists.
- Kooij, R.V. 1989. Research on children's play. *Play & Culture*, 2: 20-34.
- Lautamo, T. 2009. Assessing play in a social setting. In: K. Stagnitti & R. Cooper (Eds.), Play as therapy: Assessment and therapeutic interventions (pp. 115-129). London: Jessica Kingsley.
- Lautamo, T., Kottorp, A. & Salminen, A.-L. 2005. Play assessment for group settings: A pilot study to construct an assessment tool. *Scandinavian Journal of Occupational Therapy*, 12, 136-144.
- Lautamo, T. & Heikkilä M. 2010. Inter-rater reliability of the Play Assessment for Group Settings. *Scandinavian Journal of Occupational Therapy*, 18, 3-10.
- Lautamo, T., Laakso, M-L., Aro, T., Ahonen, T., Törmäkangas T., 2011. Validity of the Play Assessment for Group Settings: An evaluation of differential



- item functioning between children with specific language impairment and typically developing peers. *Australian Occupational Therapy Journal*, 58: 222-230: doi: 10.1111/j.1440-1630.2011.00941.x.
- Lawlor, M.C. 2003. The significance of being occupied: The social construction of childhood occupations. *American Journal of Occupational Therapy*, 57, 24-34.
- Lieberman, J.N. 1977. Playfulness: Its relationship to imagination and creativity. New York: Academic Press.
- Linacre, J.M. 1994. Sample Size and Item Calibration Stability. *Rasch Measurement Transactions*, 7:4 p.328. (Cited 5.1.2012) Available: <http://www.rasch.org/rmt/rmt74m.htm>
- Linacre, J.M. 1987-2008. FASETS, Many-faceted Rasch measurement computer program 1987-2008. Retrieved 4 September 2003, from <http://www.winsteps.com>
- Linacre JM. 2011. A user's guide to FASETS. [cited 16 March 2011]. Available at: <http://winsteps.com/winman/reliability.htm>.
- Parham, L. D. 2008. Play and occupational therapy. In: L. D. Parham & L. S. Fazio (Eds.), *Play in occupational therapy for children*, (2nd edn, pp. 3-42). St. Louis, MO: Mosby.
- Pellegrini, A.D., Dupuis, D., & Smith, P.K. 2007. Play in evolution and development. *Developmental review*, 27, 261-276.
- Schumacker, R.E. & Smith, Jr. E.V. 2007. Reliability A Rasch Perspective. *Educational and Psychological Measurement*, 67, 3, 394-409.
- Schuck, P. 2004. Quality Assessing reproducibility for interval data in health-related quality of life questionnaires: Which coefficient should be used? *Quality of Life Research*, 13: 571-586.
- Sutton-Smith, B. 1997. *The ambiguity of play*. Cambridge: Harvard.
- Stemler, S.E. 2004. A comparison of consensus, consistency, and measurement approaches to estimating inter rater reliability: Practical assessment, research and evaluation. PAREonline.net [cited 20July 2009]. Available at: <http://www.pareonline.net/getvn.asp?v = 9&n = 4>
- Tham, K., Bernspång, B., & Fisher, A.G. 1999. Development of the Assessment of Awareness of Disability. *Scandinavian Journal of Occupational Therapy*, 6, 184-90.
- Tesio, L. 2003. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *Journal of Rehabilitation Medicine*, 35, 105-115.
- Weir, J. 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, 19(1), 231-240.
- Wright, B., & Linacre, M. 1994. Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 3 [cited 30 June 2009]. Available at: <http://www.rasch.org/rmt/rmt83b.htm>

TABLE 1 Three pair of raters rated children (n=29) twice in live situations in one week time shift

<b>Time 1</b>		
Rater 1 + Rater 2 n=10	Rater 3 + Rater 4 n=10	Rater 5 + Rater 6 n=9
<b>Time 2</b>		
Rater 1 + Rater 6 n=10	Rater 3 + Rater 2 n=10	Rater 5 + Rater 4 n=9

TABLE 2 Rater calibrations (logits) and rater goodness-on-fit (MnSq, z) to the MFR model for the PAGES

<b>Rater</b>	Severity		Infit		Outfit	
	Calibration	SE	MnSq	z	MnSq	z
6	0.29	0.05	0.93	-1.3	0.95	-0.90
1	0.03	0.06	1.12	2.0	1.17	2.7
4	0.01	0.06	1.04	0.6	1.07	1.1
5	-0.09	0.06	1.16	2.9	1.12	2.1
3	-0.24	0.06	1.49	8.1	1.50	7.6

TABLE 3 Logit value difference of the individual children between two play performance measures on one week time interval.

case	Time 1			Time 2			difference T1/T2	
	raters	logit	SE	raters	logit	SE	logit	
101	1 and 2	2,54	0,53	1 and 4	0,72	0,21	1,82	**
125	2 and 5	1,16	0,26	6 alone	-0,4	0,23	1,56	**
111	3 and 4	0,3	0,22	3 and 6	1,62	0,25	1,32	**
112	3 and 4	0,36	0,22	3 and 6	1,63	0,23	1,27	**
122	5 and 6	1,08	0,21	5 and 2	-0,18	0,21	1,26	**
107	1 and 2	1,08	0,27	1 and 4	0,02	0,17	1,06	**
104	1 and 2	0,78	0,24	1 and 4	-0,11	0,18	0,89	*
124	5 and 6	0,18	0,15	5 and 2	-0,68	0,26	0,86	*
118	3 and 4	0,32	0,15	3 and 6	-0,53	0,16	0,85	*
126	2 and 5	0,82	0,24	5 and 6	1,66	0,19	0,84	*
105	1 and 2	1,11	0,35	1 and 4	0,37	0,17	0,74	*
131	5 and 6	0,65	0,17	5 and 2	-0,05	0,22	0,7	*
102	1 and 2	0,91	0,3	1 and 4	1,49	0,21	0,58	*
117	3 and 4	0,38	0,19	3 and 6	-0,19	0,18	0,57	*
108	1 and 2	0,09	0,2	1 and 4	-0,39	0,17	0,48	
115	3 and 4	0,98	0,17	3 and 6	1,46	0,21	0,48	
103	1 and 2	-0,95	0,25	1 and 4	-0,51	0,16	0,44	
114	3 and 4	1,21	0,22	3 and 6	1,65	0,2	0,44	
106	1 and 4	0,6	0,17	1 and 2	1,02	0,24	0,42	
113	3 and 4	-1,34	0,21	3 and 6	-0,96	0,16	0,38	
127	2 and 5	0,77	0,3	5 and 6	0,45	0,16	0,32	
109	1 and 2	0,02	0,21	1 and 4	-0,25	0,15	0,27	
123	5 and 6	0,4	0,17	5 and 2	0,13	0,22	0,27	
128	5 and 6	-0,26	0,28	5 and 2	-0,1	0,22	0,16	
116	1 and 3	0,53	0,2	3 and 4	0,68	0,17	0,15	
119	3 and 6	-0,22	0,2	3 and 4	-0,09	0,2	0,13	
120	3 and 6	-0,49	0,17	3 and 4	-0,62	0,18	0,13	
130	5 and 6	0,36	0,17	5 and 2	0,33	0,29	0,03	
132	5 and 6	0,06	0,15	5 and 2	0,05	0,22	0,01	

JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH

- 1 KYÖSTIÖ, O. K., Oppilaiden terveydentilan riippuvuus koulutyypistä. - Health status of pupils according to type of school. 78 p. Summary 6 p. 1962.
- 2 HEINONEN, VEIKKO, Differentiation of primary mental abilities. 136 p. 1963.
- 3 ELONEN, ANNA S., TAKALA, MARTTI & RUOPPILA ISTO, A study of intellectual functions in children by means of the KTK performance scales. 181 p. 1963.
- 4 JUURMAA, JYRKI, On the ability structure of the deaf. 114 p. 1963.
- 5 HEINONEN, VEIKKO, Lyhennetty faktori-analyysi. - A short method for factor analysis. 76 p. Summary 5 p. 1963.
- 6 PITKÄNEN, PENTTI, Fyysisen kunnon rakenne ja kehittyminen. - The structure and development of physical fitness. 163 p. Summary 10 p. 1964.
- 7 NURMI, VELI, Maamme seminaarien varsinaisen opettajakoulutuksen synty ja kehittyminen viime vuosisadalla I. - Die Entehung und Entwicklung der praktischen Lehrer-bildung unserer Lehrerseminare im vorigen Jahrhundert I. 270 p. Zusammenfassung 15 p. 1964.
- 8 NURMI, VELI, Maamme seminaarien varsinaisen opettajakoulutuksen synty ja kehittyminen viime vuosisadalla II. - Die Entstehung und Entwicklung der praktischen Lehrer-bildung unserer Lehrerseminare im vorigen Jahrhundert II. 123 p. Zusammenfassung 10 p. 1964.
- 9 NUMMENMAA, TAPIO, The language of the face. 66 p. 1964.
- 10 ISOSAARI, JUSSI, Bruno Boxström ja Sortavalan seminaarin kasvatusaineiden opetus 1882-1917. - Bruno Boxström und der Unterricht in den pädagogischen Fächern am Seminar von Sortavala 1882-1917. 167 p. Zusammenfassung II p. 1964.
- 11 NUMMENMAA, TAPIO & TAKALA, MARTTI, Parental behavior and sources of information in different social groups. 53 p. 1965.
- 12 WECKROTH, JOHAN, Studies in brain pathology and human performance I. - On the relationship between severity of brain injury and the level and structure of intellectual performance. 105 p. 1965.
- 13 PITKÄNEN, PENTTI, Ärsyke- ja reaktioanalyttisten faktorointitulosten vastaavuudesta. - On the congruence and coincidence between stimulus analytical and response analytical factor results. 223 p. Summary 14 p. 1967.
- 14 TENKKU, JUSSI, Are single moral rules absolute in Kant's ethics? 31 p. 1967.
- 15 RUOPPILA, ISTO, Nuorten ja varttuneiden opiskelijoiden väliset asenne-erot eräissä ylioppilaspohjaisissa oppilaitoksissa. - Attitude differences between young and advanced university and college students. 182 p. Summary 14 p. 1967.
- 16 KARVONEN, JUHANI, The structure, arousal and change of the attitudes of teacher education students. 118 p. 1967.
- 17 ELONEN, ANNA S., Performance scale patterns in various diagnostic groups. 53 p. 1968.
- 18 TUOMOLA, UUNO, Kansakouluntarkastajaan kohdistuvista rooli-odotuksista. - On role-expectations applied to school inspectors. 173 p. Summary 8 p. 1968.
- 19 PITKÄNEN, LEA, A descriptive model of aggression and nonaggression with applications to childrens behaviour. 208 p. 1969.
- 20 KOSKIAHO, BRIITTA, Level of living and industrialisation. 102 p. 1970.
- 21 KUUSINEN, JORMA, The meaning of another person's personality. 28 p. 1970.
- 22 VIJANEN, ERKKI, Pohjakoulutustaso ja kansakoulunopettajan kehitysympäristöjen muodostuminen. - The level of basic education in relation to the formation of the development milieus of primary school teachers. 280 s. Summary 13 p. 1970.
- 23 HAGFORS, CARL, The galvanic skin response and its application to the group registration of psychophysiological processes. 128 p. 1970.
- 24 KARVONEN, JUHANI, The enrichment of vocabulary and the basic skills of verbal communication. 47 p. 1971.
- 25 SEPPO, SIMO, Abiturienttien asenteet uskonnon-opetukseen. - The attitudes of students toward religious education in secondary school. 137 p. Summary 5 p. 1971.
- 26 RENKO MANU, Opettajan tehokkuus oppilaiden koulusaavutusten ja persoonallisuuden kehittämisessä. - Teacher's effectiveness in improving pupils' school achievements and developing their personality. 144 p. Summary 4 p. 1971.
- 27 VAHERVA, TAPIO, Koulutustulokset peruskoulun ala-asteella yhteisömuuttajien selittäminä. - Educational outcomes at the lower level of the comprehensive school in the light of ecological variables. 158 p. Summary 3 p. 1974.
- 28 OLKINUORA, ERKKI, Norm socialization. The formation of personal norms. 186 p. Tiivistelmä 4 p. 1974.
- 29 LIIKANEN, PIIRKKO, Increasing creativity through art education among pre-school children. 44 p. Tiivistelmä 4 p. 1975.
- 30 ELONEN, ANNA S., & GUYER, MELVIN, Comparison of qualitative characteristics of human figure drawings of Finnish children in various diagnostic categories. 46 p. Tiivistelmä 3 p. 1975.
- 31 KÄÄRIÄINEN, RISTO, Physical, intellectual, and personal characteristics of Down's syndrome. 114 p. Tiivistelmä 4 p. 1975.
- 32 MÄÄTTÄ, PAULA, Images of a young drug user. 112 p. Tiivistelmä 11 p. 1976.
- 33 ALANEN, PENTTI, Tieto ja demokratia. - Epistemology and democracy. 140 p. Summary 4 p. 1976.
- 34 NUPPONEN, RIITTA, Vahvistajaroolit aikuisten ja lapsen vuorovaikutuksessa. - The experimental roles of reinforcing agent in adult-child interaction. 209 p. Summary 11 p. 1977.

- 35 TEIKARI, VEIKKO, Vigilanssi-ilmiön mittaamisesta ja selitysmahdollisuuksista. – On measuring and explanation of vigilance. 163 p. Summary 2 p. 1977.
- 36 VOLANEN, RISTO, On conditions of decision making. A study of the conceptual foundations of administration. – Päätöksenteon edellytyksistä. Tutkimus hallinnon käsitteellisistä perusteista. 171 p. Tiivistelmä 7 p. 1977.
- 37 LYYTINEN, PAULA, The acquisition of Finnish morphology in early childhood. – Suomen kielen morfologisten säännönmukaisuuksien omaksuminen varhaislapsuudessa. 143 p. Tiivistelmä 6 p. 1978.
- 38 HAKAMÄKI, SIMO, Maaseudulle muutto muutto liikkeen osana. – Migration on rural areas as one element of migration as a whole. 175 p. Summary 5 p. 1978.
- 39 MOBERG, SAKARI, Leimautuminen erityispedagogiikassa. Nimikkeisiin apukoululainen ja tarkkailuluokkalainen liittyvät käsitykset ja niiden vaikutus hypoteettista oppilasta koskeviin havaintoihin. – Labelling in special education. 177 p. Summary 10 p. 1979.
- 40 AHVENAINEN, OSSI, Lukemis- ja kirjoittamishäiriöinen erityisopetuksessa. – The child with reading and writing disabilities in special education. 246 p. Summary 14 p. 1980.
- 41 HURME, HELENA, Life changes during childhood. – Lasten elämänmuutokset. 229 p. Tiivistelmä 3 p. 1981.
- 42 TUTKIMUS YHTEISKUNTAPOLITIIKAN VIITOITAJANA. Professori Leo Paukkuselle omistettu juhla kirja. 175 p. 1981.
- 43 HIRSJÄRVI, SIRKKA, Aspects of consciousness in child rearing. – Tietoisuuden ongelma kotikasvatuksessa. 259 p. 1981.
- 44 LASONEN, KARI, Siirtolaisoppilas Ruotsin kouluyhteisössä. Sosiometrinen tutkimus. – A socio-metric study of immigrant pupils in the Swedish comprehensive school. 269 p. Summary 7 p. 1981.
- 45 AJATUKSEN JA TOIMINNAN TIET. Matti Juntusen muistokirja. 274 p. 1982.
- 46 MÄKINEN, RAIMO, Teachers' work, wellbeing, and health. – Opettajan työ, hyvinvointi ja terveys. 232 p. Tiivistelmä 2 p. 1982.
- 47 KANKAINEN, MIKKO, Suomalaisen peruskoulun eriyttämisratkaisun yhteiskunnallisen taustan ja siirtymävaiheen toteutuksen arviointi. 257 p. Summary 11 p. 1982.
- 48 WALLS, GEORG, Health care and social welfare in, cooperation. 99 p. Tiivistelmä 9 p. 1982.
- 49 KOIVUKARI, MIRJAMI, Rote learning comprehension and participation by the learners in Zairian classrooms. – Mekaaninen oppiminen, ymmärtäminen ja oppilaiden osallistumisen opetukseen zairelaisissa koululuokissa. 286 p. Tiivistelmä 11 p. 1982.
- 50 KOPONEN, RITVA, An item analysis of tests in mathematics applying logistic test models. – Matematiikan kokeiden osioanalyysi logistisia testimalleja käyttäen. 187 p. Tiivistelmä 2 p. 1983.
- 51 PEKONEN, KYÖSTI, Byrokrania politiikan näkökulmasta. Poliitiikan ja byrokratian keskinäinen yhteys valtio- ja yhteiskuntaprosessin kehityksen valossa. – Bureaucracy from the viewpoint of politics. 253 p. 1983.
- 52 LYYTINEN, HEIKKI, Psychophysiology of anticipation and arousal. – Antisipaation ja viriämisen psykofysiologia. 190 p. Tiivistelmä 4 p. 1984.
- 53 KORAKIANGAS, MIKKO, Lastenneuvolan terveydenhoitajan arvioinnit viisivuotiaiden lasten psyykkisestä kehityksestä. – The psychological assessment of five-year-old children by public health centres. 227 p. Summary 14 p. 1984.
- 54 HUMAN ACTION AND PERSONALITY. Essays in honour of Martti Takala. 272 p. 1984.
- 55 MATILAINEN, JOUKO, Maanpuolustus ja eduskunta. Eduskuntaryhmien kannanotot ja koheesio maanpuolustuskysymyksissä Paasikiven-Kekkonen kaudella 1945-1978. – Defence and Parliament. 264 p. Summary 7 p. 1984.
- 56 PUOLUE, VALTIO JA EDUSTUKSELLINEN DEMOKRATIA. Pekka Nyholmille omistettu juhla kirja. – Party, state and representational democracy. 145 p. Summary 2 p. 1986.
- 57 SIISIÄINEN, MARTTI, Intressit, yhdistyslaitos ja poliittisen järjestelmän vakaisuus. – Interests, voluntary associations and the stability of the political system. 367 p. Summary 6 p. 1986.
- 58 MATTLAR, CARL-ERIK, Finnish Rorschach responses in cross-cultural context: A normative study. 166 p. Tiivistelmä 2 p. 1986.
- 59 ÄYSTÖ, SEIJA, Neuropsychological aspects of simultaneous and successive cognitive processes. – Rinnakkaisen ja peräkkäisen informaation prosessoinnin neuropsykologiasta. 205 p. Tiivistelmä 10 p. 1987.
- 60 LINDH, RAIMO, Suggestiiviset mielikuvamallit käyttäytymisen muokkaajina tarkkailuluokkalaisilla. – Suggestive covert modeling as a method with disturbed pupils. 194 p. Summary 8 p. 1987.
- 61 KORHONEN, TAPANI, Behavioral and neural short-latency and long-latency conditioned responses in the cat. – Välittömät ja viivästetyt hermostolliset ja käyttäytymisvasteet klassisen ehdollistamisen aikana kissalla. 198 p. Tiivistelmä 4 p. 1987.
- 62 PAHKINEN, TUULA, Psykoterapian vaikutus minäkäsitykseen. Psykoterapian käynnistämisen muutosprosessin vaikutus korkeakouluopiskelijoiden minäkäsitykseen. – Change in self-concept as a result of psychotherapy. 172 p. Summary 6 p. 1987.
- 63 KANGAS, ANITA, Keski-Suomen kulttuuri-toimintakokeilu tutkimuksena ja politiikkana. – The action research on cultural-activities in the Province of Central Finland. 301 p. Summary 8 p. 1988.
- 64 HURME, HELENA, Child, mother and grandmother. Intergenerational interaction in

- Finnish families. 187 p. 1988.
- 65 RASKU-PUTTONEN, HELENA, Communication between parents and children in experimental situations. - Vanhempien ja lasten kommunikointi strukturoiduissa tilanteissa. 71 p. Tiivistelmä 5 p. 1988.
- 66 TOSKALA, ANTERO, Kahvikuppineurootikkojen ja paniikkiagorafobikkojen minäkuvat minäsystemin rakenteina ja kognitiivisen oppimis-terapian perustana. - The self-images of coffee cup neurotics and panic agoraphobics as structures of a selfsystem and a basis for learning therapy. 261 p. Summary 6 p. 1988.
- 67 HAKKARAINEN, LIISA, Kuurojen yläasteen oppilaiden kirjoitetun kielen hallinta. - Mastery of written language by deaf pupils at the upper level of Comprehensive school. 281 p. Summary 11 p. 1988.
- 68 NÄTTI, JOUKO, Työmarkkinoiden lohkoutuminen. Segmentaatioteoriat, Suomen työmarkkinat ja yritysten työvoimastrategiat. - Segmentation theories, Finnish labour markets and the use of labour in retail trade. 189 p. Summary 10 p. 1989.
- 69 AALTOLA, JUHANI, Merkitys opettamisen ja oppimisen näkökulmasta Wittgensteinin myöhäisfilosofian ja pragmatismen valossa. - Meaning from the point of view of teaching and learning in the light of Wittgenstein's later philosophy and pragmatism. 249 p. Summary 6 p. 1989.
- 70 KINNUNEN, ULLA, Teacher stress over a school year. - Opettajan työstressi lukuvuoden aikana. 61 p. Tiivistelmä 3 p. 1989.
- 71 BREUER, HELMUT & RUOHO, KARI (Hrsg.), Pädagogisch-psychologische Prophylaxe bei 4-8-jährigen Kindern. - Pedagogis-psykologinen ennaltaehkäisy neljästä kahdeksaan vuoden iässä. 185 S. Tiivistelmä 1 S. 1989.
- 72 LUMMELAHTI, LEENA, Kuusivuotiaiden sopeutuminen päiväkotiin. Yksilöllistetty malliohjelma päiväkotiin heikosti sopeutuvien kuusivuotiaiden ohjauksessa sekä vanhempien kasvatuskäytännön yhteydet lapsen sopeutumiseen ja minäkäsitykseen. - The adjustment of six-year-old children to day-care-centres. 224 p. Summary 9 p. 1990.
- 73 SALOVIITA, TIMO, Adaptive behaviour of institutionalized mentally retarded persons. - Laitoksessa asuvien kehitysvammaisten adaptiivinen käyttäytyminen. 167 p. Tiivistelmä 4 p. 1990.
- 74 PALONEN, KARI et SUBRA, LEENA (Eds.), Jean-Paul Sartre - un philosophe du politique. - Jean-Paul Sartre - poliittisuuden filosofi. 107 p. Tiivistelmä 2 p. 1990.
- 75 SINIVUO, JUHANI, Kuormitus ja voimavarat upseerin uralla. - Work load and resources in the career of officers. 373 p. Summary 4 p. 1990.
- 76 PÖLKKI, PIRJO, Self-concept and social skills of school beginners. Summary and discussion. - Koulutulokkaiden minäkäsitys ja sosiaaliset taidot. 100 p. Tiivistelmä 6 p. 1990.
- 77 HUTTUNEN, JOUKO, Isän merkitys pojan sosiaaliselle sukupuolelle. - Father's impact on son's gender role identity. 246 p. Summary 9 p. 1990.
- 78 AHONEN, TIMO, Lasten motoriset koordinaatiohäiriöt. Kehitysneuropsykologinen seuranta-tutkimus. - Developmental coordination disorders in children. A developmental neuropsychological follow-up study. 188 p. Summary 9 p. 1990.
- 79 MURTO, KARI, Towards the well functioning community. The development of Anton Makarenko and Maxwell Jones' communities. - Kohti toimivaa yhteisöä. Anton Makarenkon ja Maxwell Jonesin yhteisöjen kehitys. 270 p. Tiivistelmä 5 p. Cp2`<, 5 c. 1991.
- 80 SEIKKULA, JAAKKO, Perheen ja sairaalan rajasynteesi potilaan sosiaalisessa verkostossa. - The family-hospital boundary system in the social network. 285 p. Summary 6 p. 1991.
- 81 ALANEN, ILKKA, Miten teoretisoida maa-talouden pientuotantoa. - On the conceptualization of petty production in agriculture. 360 p. Summary 9 p. 1991.
- 82 NIEMELÄ, EINO, Harjaantumisoppilas peruskoulun liikuntakasvatuksessa. - The trainable mentally retarded pupil in comprehensive school physical education. 210 p. Summary 7 p. 1991.
- 83 KARILA, IRMA, Lapsivuodeajan psyykkisten vaikeuksien ennakointi. Kognitiivinen malli. - Prediction of mental distress during puerperium. A cognitive model. 248 p. Summary 8 p. 1991.
- 84 HAAPASALO, JAANA, Psychopathy as a descriptive construct of personality among offenders. - Psykopatia rikoksentekijöiden persoonallisuutta kuvaavana konstruktiona. 73 p. Tiivistelmä 3 p. 1992.
- 85 ARNKIL, ERIK, Sosiaalityön rajasynteemit ja kehitysvyöhyke. - The systems of boundary and the developmental zone of social work. 65 p. Summary 4 p. 1992.
- 86 NIKKI, MAIJA-LIISA, Suomalaisen koulutusjärjestelmän kielikoulutus ja sen relevanssi. Osa II. - Foreign language education in the Finnish educational system and its relevance. Part 2. 204 p. Summary 5 p. 1992.
- 87 NIKKI, MAIJA-LIISA, The implementation of the Finnish national plan for foreign language teaching. - Valtakunnallisen kielenopetuksen yleissuunnitelman toimeenpano. 52 p. Yhteenveto 2 p. 1992.
- 88 VASKILAMPI, TUULA, Vaihtoehtoinen terveydenhuolto hyvinvointivaltion terveystaloudella. - Alternative medicine on the health market of welfare state. 120 p. Summary 8 p. 1992.
- 89 LAAKSO, KIRSTI, Kouluvaikeuksien ennustaminen. Käyttäytymishäiriöt ja kielelliset vaikeudet peruskoulun alku- ja päättövaiheissa. -

- Prediction of difficulties in school. 145 p. Summary 4 p. 1992.
- 90 SUUTARINEN, SAKARI, Herbartilainen pedagoginen uudistus Suomen kansakoulussa vuosisadan alussa (1900-1935). - Die Herbart'sche pädagogische Reform in den finnischen Volksschulen zu Beginn dieses Jahrhunderts (1900-1935). 273 p. Zusammenfassung 5 S. 1992.
- 91 AITTOLA, TAPIO, Uuden opiskelijatyypin synty. Opiskelijoiden elämänvaiheet ja tieteenalaspesifien habitusten muovautuminen 1980-luvun yliopistossa. - Origins of the new student type. 162 p. Summary 4 p. 1992
- 92 KORHONEN, PEKKA, The origin of the idea of the Pacific free trade area. - Tyynenmeren vapaa-kauppa-alueen idean muotoutuminen. - Taiheiyo jiyuu boeeki chi-iki koosoo no seisei. 220 p. Yhteenveto 3 p. Yooyaku 2 p. 1992.
- 93 KERÄNEN, JYRKI, Avohoitoon ja sairaalahoitoon valikoituminen perhekeskeisessä psykiatrisessa hoitojärjestelmässä. - The choice between outpatient and inpatient treatment in a family centred psychiatric treatment system. 194 p. Summary 6 p. 1992.
- 94 WAHLSTRÖM, JARL, Merkitysten muodostuminen ja muuttuminen perheterapeuttisessa keskustelussa. Diskurssianalyttinen tutkimus. - Semantic change in family therapy. 195 p. Summary 5 p. 1992.
- 95 RAHEEM, KOLAWOLE, Problems of social security and development in a developing country. A study of the indigenous systems and the colonial influence on the conventional schemes in Nigeria. - Sosiaaliturvan ja kehityksen ongelmia kehitysmaassa. 272 p. Yhteenveto 3 p. 1993.
- 96 LAINE, TIMO, Aistisuus, kehollisuus ja dialogisuus. Ludwig Feuerbachin filosofian lähtökohtia ja niiden kehitysnäkymiä 1900-luvun antropologisesti suuntautuneessa fenomenologiassa. - Sensuousness, bodilyness and dialogue. Basic principles in Ludwig Feuerbach's philosophy and their development in the anthropologically oriented phenomenology of the 1900's. 151 p. Zusammenfassung 5 S. 1993.
- 97 PENTTONEN, MARKKU, Classically conditioned lateralized head movements and bilaterally recorded cingulate cortex responses in cats. - Klassisesti ehdollistetut sivuttaiset päänliikkeet ja molemminpuoliset aivojen pihtipoimun vasteet kissalla. 74 p. Yhteenveto 3 p. 1993.
- 98 KORO, JUKKA, Aikuinen oman oppimisensa ohjaajana. Itseohjautuvuus, sen kehittyminen ja yhteys opetustuloksiin kasvatustieteen avoimen korkeakouluopetuksen monimuotokokeilussa. - Adults as managers of their own learning. Self-directiveness, its development and connection with the cognitive learning results of an experiment on distance education for the teaching of educational science. 238 p. Summary 7 p. 1993.
- 99 LAIHIALA-KANKAINEN, SIRKKA, Formaalin ja funktionaalinen traditio kieltenopetuksessa. Kieltenopetuksen oppihistoriallinen tausta antiikista valistukseen. - Formal and functional traditions in language teaching. The theory-historical background of language teaching from the classical period to the age of reason. 288 p. Summary 6 p. 1993.
- 100 MÄKINEN, TERTTU, Yksilön varhaiskehitys koulunkäynnin perustana. - Early development as a foundation for school achievement. 273 p. Summary 16 p. 1993.
- 101 KOTKAVIRTA, JUSSI, Practical philosophy and modernity. A study on the formation of Hegel's thought. - Käytännöllinen filosofia ja modernisuus. Tutkielma Hegelin ajattelun muotoutumisesta. 238 p. Zusammenfassung 3 S. Yhteenveto 3 p. 1993.
- 102 EISENHARDT, PETER L., PALONEN, KARI, SUBRA, LEENA, ZIMMERMANN RAINER E. (Eds.), Modern concepts of existentialism. Essays on Sartrean problems in philosophy, political theory and aesthetics. 168 p. Tiivistelmä 2 p. 1993.
- 103 KERÄNEN, MARJA, Modern political science and gender. A debate between the deaf and the mute. - Moderni valtio-oppi ja nainen. Mykkien ja kuuvojen välinen keskustelu. 252 p. Tiivistelmä 4 p. 1993.
- 104 MATIKAINEN, TUULA, Työtaitojen kehittyminen erityisammattikouluvaiheen aikana. - Development of working skills in special vocational school. 205 p. Summary 4 p. 1994.
- 105 PIHLAJARINNE, MARJA-LEENA, Nuoren sairastuminen skitsofreeniseen häiriöön. Perheterapeuttinen tarkastelutapa. - The onset of schizophrenic disorder at young age. Family therapeutic study. 174 p. Summary 5 p. 1994.
- 106 KUUSINEN, KIRSTI-LIISA, Psykkinen itsesäätely itsehoidon perustana. Itsehoito I-tyypin diabetesta sairastavilla aikuisilla. - Self-care based on self-regulation. Self-care in adult type I diabetics. 260 p. Summary 17 p. 1994.
- 107 MENGISTU, LEGESSE GEBRESELLASSIE, Psychological classification of students with and without handicaps. A tests of Holland's theory in Ethiopia. 209 p. 1994.
- 108 LESKINEN, MARKKU (ED.), Family in focus. New perspectives on early childhood special education. 158 p. 1994.
- 109 LESKINEN, MARKKU, Parents' causal attributions and adjustment to their child's disability. - Vanhempien syytulkinnat ja sopeutuminen lapsensa vammaisuuteen. 104 p. Tiivistelmä 1 p. 1994.
- 110 MATTHIES, AILA-LEENA, Epävirallisen sektorin ja hyvinvointivaltion suhteiden modernisoituminen. - The informal sector and the welfare state. Contemporary relationships. 63 p. Summary 12 p. 1994.
- 111 AITTOLA, HELENA, Tutkimustyön ohjaus ja ohjaussuhteet tieteellisessä jatkokoulutuksessa. - Mentoring in postgraduate education. 285 p. Summary 5 p. 1995.
- 112 LINDÉN, MIRJA, Muuttuva syövän kuva ja kokeminen. Potilaiden ja ammattilaisten tulkintoja. - The changing image and experience

- of cancer. Accounts given by patients and professionals. 234 p. Summary 5 p. 1995.
- 113 VÄLIMAA, JUSSI, Higher education cultural approach. - Korkeakoulututkimuksen kulttuurinäkökulma. 94 p. Yhteenveto 5 p. 1995.
- 114 KAIPIO, KALEVI, Yhteisöllisyys kasvatuksessa. yhteisökasvatuksen teoreettinen analyysi ja käytäntöön soveltaminen. - The community as an educator. Theoretical analysis and practice of community education. 250 p. Summary 3 p. 1995.
- 115 HÄNNIKÄINEN, MARITTA, Nukesta vauvaksi ja lapsesta lääkäriksi. Roolileikkiin siirtymisen tarkastelua piagetilaisesta ja kulttuurihistoriallisen toiminnan teorian näkökulmasta. 73 p. Summary 6 p. 1995.
- 116 IKONEN, OIVA. Adaptiivinen opetus. Oppimistutkimus harjaantumiskoulun opetussuunnitelma- ja seurantajärjestelmän kehittämisen tukena. - The adaptive teaching. 90 p. Summary 5 p. 1995.
- 117 SUUTAMA, TIMO. Coping with life events in old age. - Elämän muutos- ja ongelmatilanteiden käsittely iäkkäillä ihmisillä. 110 p. Yhteenveto 3 p. 1995.
- 118 DERSEH, TIBEBU BOGALE, Meanings Attached to Disability, Attitudes towards Disabled People, and Attitudes towards Integration. 150 p. 1995.
- 119 SAHLBERG, PASI, Kuka auttaisi opettajaa. Post-moderni näkökulma opetuksen muu-tokseen yhden kehittämisprojektin valossa. - Who would help a teacher. A post-modern perspective on change in teaching in light of a school improvement project. 255 p. Summary 4 p. 1996.
- 120 UHINKI, AILO, Distress of unemployed job-seekers described by the Zulliger Test using the Comprehensive System. - Työttömien työntekijöiden ahdinko kuvattuna Comprehensive Systemin mukaisesti käytetyillä Zulligerin testillä. 61 p. Yhteenveto 3p. 1996.
- 121 ANTIKAINEN, RISTO, Clinical course, outcome and follow-up of inpatients with borderline level disorders. - Rajatilapotilaiden osastohoidon tuloksellisuus kolmen vuoden seurantatutkimuksessa Kys:n psykiatrian klinikassa. 102 p. Yhteenveto 4 p. 1996.
- 122 RUUSUVIRTA, TIMO, Brain responses to pitch changes in an acoustic environment in cats and rabbits. - Aivovasteet kuuloärsykemuutoksiin kissoilla ja kaneilla. 45 p. Yhteenveto 2 p. 1996.
- 123 VISTI, ANNALISA, Työyhteisön ja työn tuottavuuden kehitys organisaation transformaatiossa. - Development of the work community and changes in the productivity of work during an organizational transformation process. 201 p. Summary 12 p. 1996.
- 124 SALLINEN, MIKAEL, Event-related brain potentials to changes in the acoustic environment during sleep and sleepiness. - Aivojen herätevasteet muutoksiin kuuloärsykesarjassa unen ja uneliaisuuden aikana. 104 p. Yhteenveto 3 p. 1997.
- 125 LAMMINMÄKI, TUUJA, Efficacy of a multi-faceted treatment for children with learning difficulties. - Oppimisvaikeuksien neurokognitiivisen ryhmäkuntoutuksen tuloksellisuus ja siihen vaikuttavia tekijöitä. 56 p. Yhteenveto 2 p. 1997.
- 126 LUTINEN, JAANA, Fragmentoituva kulttuuripolitiikka. Paikallisen kulttuuripolitiikan tulkintakehykset Ylä-Savossa. - Fragmenting-cultural policy. The interpretative frames of local cultural politics in Ylä-Savo. 178 p. Summary 9 p. 1997.
- 127 MARTTUNEN, MIIKA, Studying argumentation in higher education by electronic mail. - Argumentointia yliopisto-opinnoissa sähköpostilla. 60 p. (164 p.) Yhteenveto 3 p. 1997.
- 128 JAAKKOLA, HANNA, Kielitieto kielitaitoon pyrittäessä. Vieraiden kielten opettajien käsityksiä kieliopin oppimisesta ja opettamisesta. - Language knowledge and language ability. Teachers' conceptions of the role of grammar in foreign language learning and teaching. 227 p. Summary 7 p. 1997.
- 129 SUBRA, LEENA, A portrait of the political agent in Jean-Paul Sartre. Views on playing, acting, temporality and subjectivity. - Poliittisen toimijan muotokuva Jean-Paul Sartrella. Näkymiä pelaamiseen, toimintaan, ajallisuuteen ja subjektiivisuuteen. 248 p. Yhteenveto 2 p. 1997.
- 130 HAARAKANGAS, KAUKO, Hoitokokouksen äänet. Dialoginen analyysi perhekeskeisen psykiatrisen hoitoprosessin hoitokokous-keskusteluita työryhmän toiminnan näkökulmasta. - The voices in treatment meeting. A dialogical analysis of the treatment meeting conversations in family-centred psychiatric treatment process in regard to the team activity. 136 p. Summary 8 p. 1997.
- 131 MATINHEIKKI-KOKKO, KAIJA, Challenges of working in a cross-cultural environment. Principles and practice of refugee settlement in Finland. - Kulttuurienvälisen työn haasteet. Periaatteet ja käytäntö maahanmuuttajien hyvinvoinnin turvaamiseksi Suomessa. 130 p. Yhteenveto 3 p. 1997.
- 132 KIVINIEMI, KARI, Opettajuuden oppimisesta harjoittelun harhautuksiin. Aikuisopiskelijoiden kokemuksia opetusharjoittelusta ja sen ohjauksesta luokanopettajakoulutuksessa. - From the learning of teacherhood to the fabrications of practice. Adult students' experiences of teaching practice and its supervision in class teacher education. 267 p. Summary 8 p. 1997.
- 133 KANTOLA, JOUKO, Cygnaeuksen jäljillä käsityönopetuksesta teknologiseen kasvatukseen. - In the footsteps of Cygnaeus. From handicraft teaching to technological education. 211 p. Summary 7 p. 1997.
- 134 KAARTINEN, JUKKA, Nocturnal body movements



- and sleep quality. - Yölliset kehon liikkeet ja unen laatu. 85 p. Yhteenveto 3 p. 1997.
- 135 MUSTONEN, ANU, Media violence and its audience. - Mediaväkivalta ja sen yleisö. 44 p. (131 p.). Yhteenveto 2 p. 1997.
- 136 PERTTULA, JUHA, The experienced life-fabrics of young men. - Nuorten miesten koettu elämänkudelman. 218 p. Yhteenveto 3 p. 1998.
- 137 TIKKANEN, TARJA, Learning and education of older workers. Lifelong learning at the margin. - Ikääntyvän työväestön oppiminen ja koulutus. Elinikäisen oppimisen marginaalissa. 83 p. (154 p.). Yhteenveto 6 p. 1998.
- 138 LEINONEN, MARKKU, Johannes Gezelius vanhempi luonnonmukaisen pedagogiikan soveltajana. Comeniuslainen tulkinta. - Johannes Gezelius the elder as implementer of natural pedagogy. A Comenian interpretation. 237 p. Summary 7 p. 1998.
- 139 KALLIO, EEVA, Training of students' scientific reasoning skills. - Korkeakouluopiskelijoiden tieteellisen ajattelun kehittäminen. 90 p. Yhteenveto 1 p. 1998.
- 140 NIEMI-VÄKEVÄINEN, LEENA, Koulutusjaksot ja elämänpolitiikka. Koulutautuminen yksilöllistymisen ja yhteisöllisyyden risteysasemana. - Sequences of vocational education as life politics. Perspectives of individualization and communality. 210 p. Summary 6 p. 1998.
- 141 PARIKKA, MATTI, Teknologiaкомпетенssi. Teknologiakasvatuksen uudistamishaasteita peruskoulussa ja lukiossa. - Technological competence. Challenges of reforming technology education in the Finnish comprehensive and upper secondary school. 207 p. Summary 13 p. 1998.
- 142 TA OPETTAJAN APUNA - EDUCATIONAL TA FOR TEACHER. Professori Pirkko Liikaselle omistettu juhla-kirja. 207 p. Tiivistelmä - Abstract 14 p. 1998.
- 143 YLÖNEN, HILKKA, Taikahattu ja hopeakengät - sadun maailmaa. Lapsi päiväkodissa sadun kuulijana, näkijänä ja kokijana. - The world of the colden cap and silver shoes. How kinder garden children listen to, view, and experience fairy tales. 189 p. Summary 8 p. 1998.
- 144 MOILANEN, PENTTI, Opettajan toiminnan perusteiden tulkinta ja tulkinnan totuudellisuuden arviointi. - Interpreting reasons for teachers' action and the verifying the interpretations. 226 p. Summary 3p. 1998.
- 145 VAURIO, LEENA, Lexical inferencing in reading in english on the secondary level. - Sanapäättely englanninkielistä tekstiä luettaessa lukioasteella. 147 p. Yhteenveto 3 p. 1998.
- 146 ETELÄPELTO, ANNELI, The development of expertise in information systems design. - Asiantuntijuuden kehittyminen tietojärjestelmien suunnittelussa. 132 p. (221p.). Yhteenveto 12 p. 1998.
- 147 PIIRHONEN, ANTTI, Redundancy as a criterion for multimodal user-interfaces. - Käsitteistö luo näkökulman käyttöliittymäanalyyysiin. 141 p. Yhteenveto 3 p. 1998.
- 148 RÖNKÄ, ANNA, The accumulation of problems of social functioning: outer, inner, and behavioral strands. - Sosiaalinen selviytyminen lapsuudesta aikuisuuteen: ongelmien kasautumisen kolme väylää. 44 p. (129 p.) Yhteenveto 3 p. 1999.
- 149 NAUKKARINEN, AIMO, Tasapainoilua kurinalaisuuden ja tarkoituksenmukaisuuden välillä. Oppilaiden ei-toivottuun käyttäytymiseen liittyvän ongelmanratkaisun kehittäminen yhden peruskoulun yläasteen tarkastelun pohjalta. - Balancing rigor and relevance. Developing problem-solving associated with students' challenging behavior in the light of a study of an upper comprehensive school. 296 p. Summary 5 p. 1999.
- 150 HOLMA, JUHA, The search for a narrative. Investigating acute psychosis and the need-adapted treatment model from the narrative viewpoint. - Narratiivinen lähestymistapa akuuttiin psykoosiin ja tarpeenmukaisen hoidon malliin. 52 p. (105 p.) Yhteenveto 3 p. 1999.
- 151 LEPPÄNEN, PAAVO H.T., Brain responses to changes in tone and speech stimuli in infants with and without a risk for familial dyslexia. - Aivovasteet ääni- ja puheärsykkeiden muutoksiin vauvoilla, joilla on riski suvussa esiintyvään dysleksiaan ja vauvoilla ilman tätä riskiä. 100 p. (197 p.) Yhteenveto 4 p. 1999.
- 152 SUOMALA, JYRKI, Students' problem solving in the LEGO/Logo learning environment. - Oppilaiden ongelmanratkaisu LEGO/Logo oppimisympäristössä. 146 p. Yhteenveto 3 p. 1999.
- 153 HUTTUNEN, RAUNO, Opettämisen filosofia ja kritiikki. - Philosophy, teaching, and critique. Towards a critical theory of the philosophy of education. 201 p. Summary 3p. 1999.
- 154 KAREKIVI, LEENA, Ehkä en kokeilisikaan, jos .... Tutkimus ylivieskalaisten nuorten tupakoinnista ja päihteidenkäytöstä ja niihin liittyvästä terveyskasvatuksesta vuosina 1989-1998. - Maybe I wouldn't even experiment if .... A study on youth smoking and use of intoxicants in Ylivieska and related health education in 1989-1998. 256 p. Summary 4 p. 1999.
- 155 LAAKSO, MARJA-LEENA, Prelinguistic skills and early interactional context as predictors of children's language development. - Esi-kielellinen kommunikaatio ja sen vuorovaikutuksellinen konteksti lapsen kielen kehityksen ennustajana. 127 p. Yhteenveto 2 p. 1999.
- 156 MAUNO, SAIJA, Job insecurity as a psycho-social job stressor in the context of the work-family interface. - Työn epävarmuus työn psykososiaalisena stressitekijänä työn ja perheen vuorovaikutuksen kontekstissa. 59 p. (147 p.) Yhteenveto 3 p. 1999.
- 157 MÄENSIVU KIRSTI, Opettaja määrittelijänä, oppilas määriteltävänä. Sanallisen oppilaan arvioinnin sisällön analyysi. - The teacher as a determiner - the pupil to be determined -

- content analysis of the written school reports. 215 p. Summary 5 p. 1999.
- 158 FELDT, TARU, Sense of coherence. Structure, stability and health promoting role in working life. - Koherenssin rakenne, pysyvyys ja terveyttä edistävä merkitys työelämässä. 60 p. (150 p.) Yhteenveto 5 p. 2000.
- 159 MÄNTY, TARJA, Ammatillisista erityisoppilaitok-  
sista elämään. - Life after vocational special  
education. 235 p. Summary 3 p. 2000.
- 160 SARJA, ANNELI, Dialogioppiminen pienryhmäs-  
sä. Opettajaksi opiskelevien harjoitteluprosesi  
terveydenhuollon opettajankoulutuksessa. -  
Dialogic learning in a small group. The  
process of student teachers' teaching practice  
during health care education. 165 p. Summary  
7 p. 2000.
- 161 JÄRVINEN, ANITTA, Taitajat iänikuiset. - Kotkan  
ammattilukiosta valmiuksia elämään, työelä-  
mään ja jatko-opintoihin. - Age-old  
craftmasters -Kotka vocational senior  
secondary school - giving skills for life, work  
and further studies. 224 p. Summary 2 p. 2000.
- 162 KONTIO, MARJA-LIISA, Laitoksessa asuvan  
kehitysvammaisen vanhuksen haastava  
käyttäytyminen ja hoitajan käyttämiä vaiku-  
tuskeinoja. - Challenging behaviour of  
institutionalized mentally retarded elderly  
people and measures taken by nurses to  
control it. 175 p. Summary 3 p. 2000.
- 163 KILPELÄINEN, ARJA, Naiset paikkaansa etsimäs-  
sä. Aikuiskoulutus naisen elämänsä  
rakentajana. - Adult education as determinant  
of woman's life-course. 155 p. Summary 6 p.  
2000.
- 164 RIITESUO, ANNIKKI, A preterm child grows.  
Focus on speech and language during the  
first two years. - Keskonen kasvaa: puheen  
ja kielen kehitys kahtena ensimmäisenä elin-  
vuotena. 119 p. Tiivistelmä 2 p. 2000.
- 165 TAURIAINEN, LEENA, Kohti yhteistä laatua. -  
Henkilökunnan, vanhempien ja lasten laatu-  
käsitykset päiväkodin integroidussa erityis-  
ryhmässä. - Towards common quality: staff's,  
parents' and children's conceptions of quality  
in an integration group at a daycare center.  
256 p. Summary 6 p. 2000.
- 166 RAUDASKOSKI, LEENA, Ammattikorkeakoulun  
toimintaperustaa etsimässä. Toimilupahake-  
musten sisällönanalyttinen tarkastelu. - In  
search for the founding principles of the  
Finnishpolytechnic institutes. A content  
analysis of the licence applications. 193 p.  
Summary 4 p. 2000.
- 167 TAKKINEN, SANNA, Meaning in life and its  
relation to functioning in old age. - Elämän  
tarkoituksellisuus ja sen yhteydet toiminta-  
kykyyn vanhuudessa. 51 p. (130 p.)  
Yhteenveto 2 p. 2000.
- 168 LAUNONEN, LEEVI, Eettinen kasvatusajattelu  
suomalaisen koulun pedagogisissa teksteissä  
1860-luvulta 1990-luvulle. - Ethical thinking  
in Finnish school's pedagogical texts from the  
1860s to the 1990s. 366 p. Summary 3 p. 2000.
- 169 KUORELAHTI, MATTI, Sopeutumattomien luokka-  
muotoisen erityisopetuksen tuloksellisuus. -  
The educational outcomes of special classes  
for emotionally/ behaviorally disordered  
children and youth. 176 p. Summary 2p.  
2000.
- 170 KURUNMÄKI, JUSSI, Representation, nation and  
time. The political rhetoric of the 1866  
parliamentary reform in Sweden. - Edustus,  
kansakunta ja aika. Poliittinen retoriikka  
Ruotsin vuoden 1866 valtiopäiväreformissa.  
253 p. Tiivistelmä 4 p. 2000.
- 171 RASINEN, AKI, Developing technology  
education. In search of curriculum elements  
for Finnish general education schools. 158 p.  
Yhteenveto 2 p. 2000.
- 172 SUNDHOLM, LARS, Itseohjautuvuus organisaatio-  
muutoksessa. - Self-determination in  
organisational change. 180 p. Summary 15 p.  
2000.
- 173 AHONNISKKA-ASSA, JAANA, Analyzing change in  
repeated neuropsychological assessment. 68  
p. (124 p.) Yhteenveto 2 p. 2000.
- 174 HOFFRÉN, JARI, Demokraattinen eetos - rajoista  
mahdollisuuksiin. - The democratic ethos.  
From limits to possibilities? 217 p. Summary  
2 p. 2000.
- 175 HEIKKINEN, HANNU L. T., Toimintatutkimus,  
tarinat ja opettajaksi tuleminen taito.  
Narratiivisen identiteettityön kehittäminen  
opettajankoulutuksessa toimintatutkimuksen  
avulla. - Action research, narratives and the  
art of becoming a teacher. Developing  
narrative identity work in teacher education  
through action research. 237 p. Summary 4 p.  
2001.
- 176 VUORENMAA, MARITTA, Ikkunoita arvioin-  
nin tuolle puolen. Uusia avauksia suoma-  
laiseen koulutusta koskevaan evaluaatio-  
keskusteluun. - Views across assessment:  
New openings into the evaluation  
discussion on Finnish education. 266 p.  
Summary 4 p. 2001.
- 177 LITMANEN, TAPIO, The struggle over risk. The  
spatial, temporal, and cultural dimensions of  
protest against nuclear technology. - Kamp-  
pailu riskistä. Ydinteknologian vastaisen  
protestin tilalliset, ajalliset ja kulttuuriset  
ulottuvuudet. 72 p. (153 p.) Yhteenveto 9 p.  
2001.
- 178 AUNOLA, KAISA, Children's and adolescents'  
achievement strategies, school adjustment,  
and family environment. - Lasten ja nuorten  
suoritusstrategiat koulu- ja perheympäristöis-  
sä. 51 p. (153 p.) Yhteenveto 2 p. 2001.
- 179 OKSANEN, ELINA, Arvioinnin kehittäminen  
erityisopetuksessa. Diagnosoinnista oppimi-  
sen ohjaukseen laadullisena tapaustutkimuk-  
sena. - Developing assessment practices in  
special education. From a static approach to

- dynamic approach applying qualitative case. 182 p. Summary 5 p. 2001.
- 180 VIITTALA, KAISU, "Kyllä se tommosellaki lapsella on kovempi urakka". Sikiöaikana alkoholille altistuneiden huostaanotettujen lasten elämäntilanne, riskiprosessit ja suojaavat prosessit. - "It's harder for that kind of child to get along". The life situation of the children exposed to alcohol in utero and taken care of by society, their risk and protective processes. 316 p. Summary 4 p. 2001.
- 181 HANSSON, LEENI, Networks matter. The role of informal social networks in the period of socioeconomic reforms of the 1990s in Estonia. - Verkostoilla on merkitystä: infor-maalisten sosiaalisten verkostojen asema Virossa 1990-luvun sosio-ekonomisten muutosten aikana. 194 p. Yhteenveto 3 p. 2001.
- 182 BÖÖK, MARJA LEENA, Vanhemmuus ja vanhemmuuden diskurssit työttömyystilanteessa. - Parenthood and parenting discourses in a situation of unemployment. 157 p. Summary 5 p. 2001.
- 183 KOKKO, KATJA, Antecedents and consequences of long-term unemployment. - Pitkäaikaistyöttömyyden ennakoijia ja seurauksia. 53 p. (115 p.) Tiivistelmä 3 p. 2001.
- 184 KOKKONEN, MARJA, Emotion regulation and physical health in adulthood: A longitudinal, personality-oriented approach. - Aikuisiän tunteiden säätely ja fyysinen terveys: pitkittäistutkimuksellinen ja persoonallisuuskeskeinen lähestymistapa. 52 p. (137 p.) Tiivistelmä 3 p. 2001.
- 185 MÄNNIKKÖ, KAISA, Adult attachment styles: A Person-oriented approach. - Aikuisten kiintymystyyli. 142 p. Yhteenveto 5 p. 2001.
- 186 KATVALA, SATU, Missä äiti on? Äitejä ja äitiyden uskomuksia sukupolvien saatossa. - Where's mother? Mothers and maternal beliefs over generations. 126 p. Summary 3 p. 2001.
- 187 KIISKINEN, ANNA-LIISA, Ympäristöhallinto vastuullisen elämäntavan edistäjänä. - Environmental administration as promoter of responsible living. 229 p. Summary 8 p. 2001.
- 188 SIMOLA, AHTI, Työterveyshuolto-organisaation toiminta, sen henkilöstön henkinen hyvinvointi ja toiminnan tuloksellisuus. - Functioning of an occupational health service organization and its relationship to the mental well-being of its personnel, client satisfaction, and economic profitability. 192 p. Summary 12 p. 2001.
- 189 VESTERINEN, PIRKKO, Projektiopiskelu- ja oppiminen ammattikorkeakoulussa. - Project-based studying and learning in the polytechnic. 257 p. Summary 5 p. 2001.
- 190 KEMPPAINEN, JAANA, Kotikasvatus kolmessa sukupolvessa. - Childrearing in three generations. 183 p. Summary 3 p. 2001.
- 191 HOHENTHAL-ANTIN LEONIE, Luvan ottaminen - Ikäihmiset teatterin tekijöinä. - Taking permission- Elderly people as theatre makers. 183 p. Summary 5 p. 2001.
- 192 KAKKORI, LEENA, Heideggerin aukeama. Tutkimuksia totuudesta ja taiteesta Martin Heideggerin avaamassa horisontissa. - Heidegger's clearing. Studies on truth and art in the horizon opened by Martin Heidegger. 156 p. Summary 2 p. 2001.
- 193 NÄRHI, VESA, The use of clinical neuropsychological data in learning disability research. - Asiakastyön yhteydessä kerätyn neuropsykologisen aineiston käyttö oppimisvaikeustutkimuksessa. 103 p. Yhteenveto 2 p. 2002.
- 194 SUOMI, ASTA, Ammattia etsimässä. Aikuisopiskelijat kertovat sosiaaliohjaajakoulutuksesta ja narratiivisen pätevyyden kehittymisestä. - Searching for professional identity. Adult students' narratives on the education of a social welfare supervisor and the development of narrative competence. 183 p. Summary 2 p. 2002.
- 195 PERKKILÄ, PÄIVI, Opettajien matematiikkauskomukset ja matematiikan oppikirjan merkitys alkuopetuksessa. 212 p. - Teacher's mathematics beliefs and meaning of mathematics textbooks in the first and the second grade in primary school. Summary 2 p. 2002.
- 196 VESTERINEN, MARJA-LIISA, Ammatillinen harjoittelu osana asiantuntijuuden kehittymistä ammattikorkeakoulussa. - Promoting professional expertise by developing practical learning at the polytechnic. 261 p. Summary 5 p. 2002.
- 197 POHJANEN, JORMA, Mitä kello on? Kello modernissa yhteiskunnassa ja sen sosiologisessa teoriassa. - What's the time. Clock on modern society and in it's sociological theory. 226 p. Summary 3 p. 2002.
- 198 RANTALA, ANJA, Perhekeskeisyys - puhetta vai todellisuutta? Työntekijöiden käsitykset yhteistyöstä erityistä tukea tarvitsevan lapsen perheen kanssa. - Family-centeredness rhetoric or reality? Summary 3 p. 2002.
- 199 VALANNE, EIJA, "Meidän lapsi on arvokas" Henkilökohtainen opetuksen järjestämistä koskeva suunnitelma (HOJKS) kunnallisessa erityiskoulussa. - "Our child is precious" - The individual educational plan in the context of the special school. 219 p. Yhteenveto 2 p. 2002.
- 200 HOLOPAINEN, LEENA, Development in reading and reading related skills; a follow-up study from pre-school to the fourth grade. 57 p. (138 p.) Yhteenveto 3 p. 2002.
- 201 HEIKKINEN, HANNU, Draaman maailmat oppimisalueina. Draamakasvatuksen vakava leikillisuus. - Drama worlds as learning areas - the serious playfulness or drama education. 164 p. Summary 5 p. 2002.
- 202 HYTÖNEN, TUUJA, Exploring the practice of human resource development as a field of professional expertise. - Henkilöstön

- kehittämistyön asiantuntijuuden rakentumisen. 137 p. (300 p.) Yhteenveto 10 p. 2002.
- 203 RIPATTI, MIKKO, Arvid Järnefeldt kasvatusajattelijana. 246 p. Summary 4 p. 2002.
- 204 VIRMASALO, ILKKA, Perhe, työttömyys ja lama. - Families, unemployment and the economic depression. 121 p. Summary 2 p. 2002.
- 205 WIKGREN, JAN, Diffuse and discrete associations in aversive classical conditioning. - Täsmälliset ja laaja-alaiset ehdollistumat klassisessa aversiivisessä ehdollistumisessa. 40 p. (81 p.) Yhteenveto 2 p. 2002.
- 206 JOKIVUORI, PERTTI, Sitoutuminen työorganisaatioon ja ammattijärjestöön. - Kilpailevia vai täydentäviä? - Commitment to organisation and trade union. Competing or complementary? 132 p. Summary 8 p. 2002.
- 207 GONZÁLEZ VEGA, NARCISO, Factors affecting simulator-training effectiveness. 162 p. Yhteenveto 1 p. 2002.
- 208 SALO, KARI, Teacher Stress as a Longitudinal Process - Opettajien stressiprosessi. 67 p. Yhteenveto 2 p. 2002.
- 209 VAUHKONEN, JOUNI, A rhetoric of reduction. Bertrand de Jouvenel's pure theory of politics as persuasion. 156 p. Tiivistelmä 2 p. 2002.
- 210 KONTONIEMI, MARITA, "Milloin sinä otat itseäsi niskasta kiinni?" Opettajien kokemuksia alisuoriutujista. - "When will you pull your socks up?" Teachers' experiences of underachievers. 218 p. Summary 3 p. 2003.
- 211 SAUKKONEN, SAKARI, Koulu ja yksilöllisyys; Jännitteitä, haasteita ja mahdollisuuksia. - School and individuality: Tensions, challenges and possibilities. 125 p. Summary 3 p. 2003.
- 212 VILJAMAA, MARJA-LEENA, Neuvola tänään ja huomenna. Vanhemmuuden tukeminen, perhekeskeisyys ja vertaistuki. - Child and maternity welfare clinics today and tomorrow. Supporting parenthood, family-centered services and peer groups. 141 p. Summary 4 p. 2003.
- 213 REMES, LIISA, Yrittäjyyskasvatuksen kolme diskurssia. - Three discourses in entrepreneurial learning. 204 p. Summary 2 p. 2003.
- 214 KARJALA, KALLE, Neulanreiästä panoraamaksi. Ruotsin kulttuurikuvan ainekset eräissä keskikoulun ja B-ruotsin vuosina 1961-2002 painetuissa oppikirjoissa. - From pinhole to panorama - The culture of Sweden presented in some middle and comprehensive school textbooks printed between 1961 and 2002. 308 p. Summary 2 p. 2003.
- 215 LALLUKKA, KIRSI, Lapsuusikä ja ikä lapsuudessa. Tutkimus 6-12-vuotiaiden sosiokulttuurisesta ikätiedosta. - Childhood age and age in childhood. A study on the sociocultural knowledge of age. 234 p. Summary 2 p. 2003.
- 216 PUUKARI, SAULL, Video Programmes as Learning Tools. Teaching the Gas Laws and Behaviour of Gases in Finnish and Canadian Senior High Schools. 361 p. Yhteenveto 6 p. 2003.
- 217 LOISA, RAIJA-LEENA, The polysemous contemporary concept. The rhetoric of the cultural industry. - Monimerkityksinen nykykäsite. Kulttuuriteollisuuden retoriikka. 244 p. Yhteenveto 2 p. 2003.
- 218 HOLOPAINEN, ESKO, Kuullun ja luetun tekstin ymmärtämisstrategiat ja -vaikeudet peruskoulun kolmannella ja yhdeksännellä luokalla. - Strategies for listening and reading comprehension and problematic listening and reading comprehension of the text during the third and ninth grades of primary school. 135 p. Summary 3 p. 2003.
- 219 PENTTINEN, SEPPÖ, Lähtökohdat liikuntaa opettavaksi luokanopettajaksi. Nuoruuden kasvuympäristöt ja opettajankoulutus opettajuuden kehitystekijöinä. - Starting points for a primary school physical education teacher. The growth environment of adolescence and teacher education as developmental factors of teachership. 201 p. Summary 10 p. 2003.
- 220 IKÄHEIMO, HEIKKI, Tunnustus, subjektiviteetti ja inhimillinen elämänmuoto: Tutkimuksia Hegelistä ja persoonien välisistä tunnustus-suhteista. - Recognition, subjectivity and the human life form: studies on Hegel and interpersonal recognition. 191 p. Summary 3 p. 2003.
- 221 ASUNTA, TUULA, Knowledge of environmental issues. Where pupils acquire information and how it affects their attitudes, opinions, and laboratory behaviour - Ympäristöasioita koskeva tieto. Mistä oppilaat saavat informaatiota ja miten se vaikuttaa heidän asenteisiinsa, mielipiteisiinsä ja laboratoriokäyttäytymiseensä. 159 p. Yhteenveto 4 p. 2003.
- 222 KUJALA, ERKKI, Sodan pojat. Sodanaikaisten pikkupoikien lapsuuskokemuksia isyyden näkökulmasta - The sons of war. 229 p. Summary 2 p. 2003.
- 223 JUSSI KURUNMÄKI & KARI PALOINEN (Hg./eds.) Zeit, Geschichte und Politik. Time, history and politics. *Zum achtzigsten Geburtstag von Reinhart Koselleck*. 310 p. 2003.
- 224 LAITINEN, ARTO, Strong evaluation without sources. On Charles Taylor's philosophical anthropology and cultural moral realism. - Vahvoja arvostuksia ilman lähteitä. Charles Taylorin filosofisesta antropologiasta ja kulturalistisesta moraalirealismista. 358 p. Yhteenveto 4 p. 2003.
- 225 GUTTORM, TOMI K. Newborn brain responses measuring feature and change detection and predicting later language development in children with and without familial risk for dyslexia. - Vastasyntyneiden aivovasteet puheäänteiden ja niiden muutosten havaitsemisessa sekä myöhemmän kielen kehityksen ennustamisessa dysleksia-riskilapsilla. 81 p. (161 p.) Yhteenveto 3 p. 2003.

- 226 NAKARI, MAIJA-LIISA, Työilmapiiri, työntekijöiden hyvinvointi ja muutoksen mahdollisuus - Work climate, employees' well-being and the possibility of change. 255 p. Summary 3 p. 2003.
- 227 METSÄPELTO, RIITTA-LEENA, Individual differences in parenting: The five-factor model of personality as an explanatory framework - Lastenkasvatus ja sen yhteys vanhemman persoonallisuuden piirteisiin. 53 p. (119 p.) Tiivistelmä 3 p. 2003.
- 228 PULKKINEN, OILI, The labyrinth of politics - A conceptual approach to the modes of the political in the scottish enlightenment. 144 p. Yhteenveto 2 p. 2003.
- 229 JUUJARVI, PETRI, A three-level analysis of reactive aggression among children. - Lasten aggressiivisiin puolustusreaktioihin vaikuttavien tekijöiden kolmitasoinen analyysi. 39 p. (115 p.) Yhteenveto 2 p. 2003.
- 230 POIKONEN, PIIRJO-LIISA, "Opetussuunnitelma on sitä elämää". Päiväkoti-koulu yhteisö opetussuunnitelman kehittäjänä. - "The curriculum is part of our life". The day-care - primary school community as a curriculum developer. 154 p. Summary 3 p. 2003.
- 231 SOININEN, SUVI, From a 'Necessary Evil' to an art of contingency: Michael Oakeshott's conception of political activity in British postwar political thought. 174 p. Summary 2p. 2003.
- 232 ALARAUDANJOKI, ESA, Nepalese child labourers' life-contexts, cognitive skills and well-being. - Työssäkäyvien nepalilaislasten elämänkonteksti, kognitiiviset taidot ja hyvinvointi. 62 p. (131 p.) Yhteenveto 4 p. 2003.
- 233 LERKKANEN, MARJA-KRISTINA, Learning to read. Reciprocal processes and individual pathways. - Lukemaan oppiminen: vastavuoroiset prosessit ja yksilölliset oppimispolut. 70 p. (155 p.) Yhteenveto 5 p. 2003.
- 234 FRIMAN, MERVI, Ammatillisen asiantuntijan etiikka ammattikorkeakoulutuksessa. - The ethics of a professional expert in the context of polytechnics. 199 p. 2004.
- 235 MERONEN, AULI, Viittomakielen omaksumisen yksilölliset tekijät. - Individual differences in sign language abilities. 110 p. Summary 5 p. 2004.
- 236 TIILIKKALA, LIISA, Mestarista tuutoriksi. Suomalaisen ammatillisen opettajuuden muutos ja jatkuvuus. - From master to tutor. Change and continuity in Finnish vocational teacherhood. 281 p. Summary 3 p. 2004.
- 237 ARO, MIKKO, Learning to read: The effect of orthography. - Kirjoitusjärjestelmän vaikutus lukemaan oppimiseen. 44 p. (122 p.) Tiivistelmä 2 p. 2004.
- 238 LAAKSO, ERKKI, Draamakokemusten äärellä. Prosessidraaman oppimispotentiaali opettajaksi opiskelevien kokemusten valossa. - Encountering drama experiences. The learning potential of process drama in the light of student teachers' experiences. 230 p. Summary 7 p. 2004.
- 239 PERÄLÄ-LITTUNEN, SATU, Cultural images of a good mother and a good father in three generations. - Kulttuuriset mielikuvat hyvästä äidistä ja hyvästä isästä kolmessa sukupolvessa. 234 p. Yhteenveto 7 p. 2004.
- 240 RINNE-KOISTINEN, EVA-MARITA, Perceptions of health: Water and sanitation problems in rural and urban communities in Nigeria. 129 p. (198 p.) Yhteenveto 3 p. 2004.
- 241 PALMROTH, AINO, Käännösten kautta kollektiiviin. Tuuliosuuskunnat toimijaverkkoina. - From translation to collective. Wind turbine cooperatives as actor networks. 177 p. Summary 7 p. 2004.
- 242 VIERIKKO, ELINA, Genetic and environmental effects on aggression. - Geneettiset ja ympäristötekijät aggressiivisuudessa. 46 p. (108 p.) Tiivistelmä 3 p. 2004.
- 243 NÄRHI, KATI, The eco-social approach in social work and the challenges to the expertise of social work. - Ekososiaalinen viitekehys ja haasteet sosiaalityön asiantuntijuudelle. 106 p. (236 p.) Yhteenveto 7 p. 2004.
- 244 URSIN, JANI, Characteristics of Finnish medical and engineering research group work. - Tutkimusryhmätyöskentelyn piirteet lääke- ja teknisissä tieteissä. 202 p. Yhteenveto 9 p. 2004.
- 245 TREUTHARDT, LEENA, Tulostehokkuuden yhteiskunnallisuus Jyväskylän yliopistossa. Tarkastelunäkökulmina muoti ja seurustelu. - The management by results a fashion and social interaction at the University of Jyväskylä. 228 p. Summary 3 p. 2004.
- 246 MATTHIES, JÜRGEN, Umweltpädagogik in der Postmoderne. Eine philosophische Studie über die Krise des Subjekts im umweltpädagogischen Diskurs. - Ympäristökasvatus postmodernissa. Filosofinen tutkimus subjektin kriisistä ympäristökasvatuksen diskurssissa. 400 p. Yhteenveto 7 p. 2004.
- 247 LAITILA, AARNO, Dimensions of expertise in family therapeutic process. - Asiantuntijuuden ulottuvuuksia perheterapeuttisessa prosessissa. 54 p. (106 p.) Yhteenveto 3 p. 2004.
- 248 LAAMANEN (ASTIKAINEN), PIIA, Pre-attentive detection of changes in serially presented stimuli in rabbits and humans. - Muutoksen esitietoinen havaitseminen sarjallisesti esitetyissä ärsykkeissä kaneilla ja ihmisillä. 35 p. (54 p.) Yhteenveto 3 p. 2004.
- 249 JUUSENAHO, RIITTA, Peruskoulun rehtoreiden johtamisen eroja. Sukupuolinen näkökulma. - Differences in comprehensive school leadership and management. A gender-based approach. 176p. Summary 3 p. 2004.

- 250 VAARAKALLIO, TUULA, "Rotten to the Core". Variations of French nationalist anti-system rhetoric. - "Systeemi on mätä". Ranskalaisien nationalistien järjestelmän vastainen retoriikka. 194 p. Yhteenveto 3 p. 2004.
- 251 KUUSINEN, PATRIK, Pitkäaikainen kipu ja depressio. Yhteyttä säätelevät tekijät. - Chronic pain and depression: psychosocial determinants regulating the relationship. 139 p. Summary 8 p. 2004.
- 252 HÄNNIKÄINEN-UUTELA, ANNA-LIISA, Uudelleen juurtuneet. Yhteisökasvatus vaikeasti päihderiippuvaisten narkomaanien kuntoutuksessa. - Rooted again. Community education in the rehabilitation of substance addicts. 286 p. Summary 3 p. 2004.
- 253 PALONIEMI, SUSANNA, Ikä, kokemus ja osaaminen työelämässä. Työntekijöiden käsityksiä iän ja kokemuksen merkityksestä ammatillisessa osaamisessa ja sen kehittämisessä. - Age, experience and competence in working life. Employees' conceptions of the the meaning and experience in professional competence and its development. 184 p. Summary 5 p. 2004.
- 254 RUIZ CEREZO, MONTSE, Anger and Optimal Performance in Karate. An Application of the IZOF Model. 55 p. (130 p.) Tiivistelmä 2 p. 2004.
- 255 LADONLAHTI, TARJA, Haasteita palvelujärjestelmälle. Kehitysvammaiseksi luokiteltu henkilö psykiatrisessa sairaalassa. - Challenges for the human service system. Living in a psychiatric hospital under the label of mental retardation. 176 p. Summary 3 p. 2004.
- 256 KOVANEN PÄIVI, Oppiminen ja asiantuntijuus varhaiskasvatuksessa. Varhaisen oppimaan ohjaamisen suunnitelma erityistä tukea tarvitsevien lasten ohjauksessa. - Learning and expertise in early childhood education. A pilot work in using VARSU with children with special needs. 175 p. Summary 2 p. 2004.
- 257 VILMI, VEIKKO, Turvallinen koulu. Suomalaisien näkemyksiä koulutuspalvelujen kansallisesta ja kunnallisesta priorisoinnista. - Secure education. Finnish views on the national and municipal priorities of Finland's education services. 134 p. Summary 5 p. 2005.
- 258 ANTTILA, TIMO, Reduced working hours. Reshaping the duration, timing and tempo of work. 168 p. Tiivistelmä 2 p. 2005.
- 259 UGASTE, AINO, The child's play world at home and the mother's role in the play. 207 p. Tiivistelmä 5 p. 2005.
- 260 KURRI, KATJA, The invisible moral order: Agency, accountability and responsibility in therapy talk. 38 p. (103 p.). Tiivistelmä 1 p. 2005.
- 261 COLLIN, KAIJA, Experience and shared practice - Design engineers' learning at work. - Suunnitteluinsinöörien työssä oppiminen - kokemuksellisuutta ja jaettuja käytäntöjä. 124 p. (211 p.). Yhteenveto 6 p. 2005.
- 262 KURKI, EIJA, Näkyvä ja näkymätön. Nainen Suomen helluntailiikkeen kentällä. - Visible and invisible. Women in the Finnish pentecostal movement. 180 p. Summary 2 p. 2005.
- 263 HEIMONEN, SIRKKALIISA, Työikäisenä Alzheimerin tautiin sairastuneiden ja heidän puolisoitensa kokemukset sairauden alkuvaiheessa. - Experiences of persons with early onset Alzheimer's disease and their spouses in the early stage of the disease. 138 p. Summary 3 p. 2005.
- 264 PIIRÖINEN, HANNU, Epävarmuus, muutos ja ammatilliset jännitteet. Suomalainen sosiaalityö 1990-luvulla sosiaalityöntekijöiden tulkinnoissa. - Uncertainty, change and professional tensions. The Finnish social work in the 1990s in the light of social workers' representations. 207 p. Summary 2 p. 2005.
- 265 MÄKINEN, JARMO, Säätiö ja maakunta. Maakuntarahastojärjestelmän kentät ja verkostot. - Foundation and region: Fields and networks of the system of the regional funds. 235 p. Summary 3 p. 2005.
- 266 PETRELIUS, PÄIVI, Sukupuoli ja subjektiivinen sosiaalityössä. Tulkintoja naistyöntekijöiden muistoista. - Gender and subjectivity in social work - interpreting women workers' memories. 67 p. (175 p.) 2005.
- 267 HOKKANEN, TIINA, Äitinä ja isänä eron jälkeen. Yhteishuoltajavanhemmuus arjen kokemuksena. - As a mother and a father after divorce. Joint custody parenthood as an everyday life experience. 201 p. Summary 8 p. 2005.
- 268 HANNU SIRKKILÄ, Elättäjyyttä vai erotiikkaa. Miten suomalaiset miehet legitimoivat parisuhteensa thaimaalaisen naisen kanssa? - Breadwinner or eroticism. How Finnish men legitimize their partnerships with Thai women. 252 p. Summary 4 p. 2005.
- 269 PENTTINEN, LEENA, Gradupuhetta tutkielmaseminaarissa. - Thesis discourse in an undergraduate research seminar. 176 p. Summary 8 p. 2005.
- 270 KARVONEN, PIIRKKO, Päiväkotilasten lukuleikit. Lukutaidon ja lukemistietoisuuden kehittyminen interventiotutkimuksessa- Reading Games for Children in Daycare Centers. The Development of Reading Ability and Reading Awareness in an Intervention Study . 179 p. Summary 3 p. 2005.
- 271 KOSONEN, PEKKA A., Sosiaalialan ja hoitotyön asiantuntijuuden kehitysehdot ja opiskelijavalinta. - Conditions of expertise development in nursing and and social care, and criteria for student selection. 276 p. Summary 3 p. 2005.

- 272 NIIRANEN-LINKAMA, PÄIVI, Sosiaalisen transformaatio sosiaalialan asiantuntijuu- den diskurssissa. - Transformation of the social in the discourse of social work expertise. 200 p. Summary 3 p. 2005.
- 273 KALLA, OUTI, Characteristics, course and outcome in first-episode psychosis. A cross-cultural comparison of Finnish and Spanish patient groups. - Ensiker- talaisten psykoosipotilaiden psyykkis- sosiaaliset ominaisuudet, sairaudenkulku ja ennuste. Suomalaisten ja espanjalaisten potilasryhmien vertailu. 75 p. (147 p.) Tiivistelmä 4 p. 2005.
- 274 LEHTOMÄKI, ELINA, Pois oppimisyhteiskun- nan marginaalista? Koulutuksen merkitys vuosina 1960-1990 opiskelleiden lapsuu- desta kuurojen ja huonokuuloisten aikuisten elämänculussa. - Out from the margins of the learning society? The meaning of education in the life course of adults who studied during the years 1960-1990 and were deaf or hard-of-hearing from childhood. 151 p. Summary 5 p. 2005.
- 275 KINNUNEN, MARJA-LIISA, Allostatic load in relation to psychosocial stressors and health. - Allostaattinen kuorma ja sen suhde psykososiaalisiin stressitekijöihin ja terveyteen. 59 p. (102 p.) Tiivistelmä 3 p. 2005.
- 276 UOINEN, VIRPI, I'm as old as I feel. Subjective age in Finnish adults. - Olen sen ikäinen kuin tunnen olevani. Suomalaisten aikuis- ten subjektiivinen ikä. 64 p. (124 p.) Tiivistelmä 3 p. 2005.
- 277 SALOKOSKI, TARJA, Tietokonepelit ja niiden pelaaminen. - Electronic games: content and playing activity. 116 p. Summary 5 p. 2005.
- 278 HIIHNALA, KAUKO, Laskutehtävien suoritta- misesta käsitteiden ymmärtämiseen. Perus- koululaisen matemaattisen ajattelun kehittyminen aritmetiikasta algebraan siirryttäessä. - Transition from the performing of arithmetic tasks to the understanding of concepts. The development of pupils' mathematical thinking when shifting from arithmetic to algebra in comprehensive school. 169 p. Summary 3 p. 2005.
- 279 WALLIN, RISTO, Yhdistyneet kansakunnat organisaationa. Tutkimus käsitteellisestä muutoksesta maailmanjärjestön organi- soinnin periaatteissa - From the league to UN. The move to an organizational vocabulary of international relations. 172 p. Summary 2 p. 2005.
- 280 VALLEALA, ULLA MAIJA, Yhteinen ymmär- täminen koulutuksessa ja työssä. Kontekstin merkitys ymmärtämisessä opiskelijaryh- män ja työtiimin keskusteluissa. - Shared understanding in education and work. Context of understanding in student group and work team discussions. 236 p. Summary 7 p. 2006.
- 281 RASINEN, TUUJA, Näkökulmia vieraskieliseen perusopetukseen. Koulun kehittämishank- keesta koulun toimintakulttuuriksi. - Perspectives on content and language integrated learning. The impact of a development project on a school's activities. 204 . Summary 6 p. 2006.
- 282 VIHOLAINEN, HELENA, Suvussa esiintyvän lukemsvaikeusriskin yhteys motoriseen ja kielelliseen kehitykseen. Tallaako lapsi kielensä päälle? - Early motor and language development in children at risk for familial dyslexia. 50 p. (94 p.) Summary 2 p. 2006.
- 283 KIILLI, JOHANNA, Lasten osallistumisen voimavarat. Tutkimus Ipanoiden osallistu- misesta. - Resources for children's participation. 226 p. Summary 3 p. 2006.
- 284 LEPPÄMÄKI, LAURA, Tekijänoikeuden oikeut- taminen. - The justification of copyright. 125 p. Summary 2 p. 2006.
- 285 SANAKSENAHO, SANNA, Eriarvoisuus ja luottamus 2000-luvun taitteen Suomessa. Bourdieulainen näkökulma. - Inequality and trust in Finland at the turn of the 21st century: Bourdieuan approach. 150 p. Summary 3 p. 2006.
- 286 VALKONEN, LEENA, Millainen on hyvä äiti tai isä? Viides- ja kuudesluokkalaisten lasten vanhemmuuskäsitykset. - What is a good father or good mother like? Fifth and sixth graders' conceptions of parenthood. 126 p. Summary 5 p. 2006.
- 287 MARTIKAINEN, LIISA, Suomalaisten nuorten aikuisten elämään tyytyväisyyden monet kasvot. - The many faces of life satisfaction among Finnish young adults. 141 p. Summary 3 p. 2006.
- 288 HAMARUS, PÄIVI, Koulukiusaaminen ilmiönä. Yläkoulun oppilaiden kokemuksia kiusaamisesta. - School bullying as a phenomenon. Some experiences of Finnish lower secondary school pupils. 265 p. Summary 6 p. 2006.
- 289 LEPPÄNEN, ULLA, Development of literacy in kindergarten and primary school. Tiivistelmä 2 p. 49 p. ( 145 p.) 2006.
- 290 KORVELA, PAUL-ERIK, The Machiavellian reformation. An essay in political theory. 171 p. Tiivistelmä 2 p. 2006.
- 291 METSOMÄKI, MARJO, "Suu on syömistä varten". Lasten ja aikuisten kohtaamisia ryhmäperhepäiväkodin ruokailutilanteissa. - Encounters between children and adults in group family day care dining situations. 251 p. Summary 3 p. 2006.
- 292 LATVALA, JUHA-MATTI, Digitaalisen kommuni- kaatiosovelluksen kehittäminen kodin ja koulun vuorovaikutuksen edistämiseksi. - Development of a digital communication system to facilitate interaction between home and school. 158 p. Summary 7 p. 2006.

- 293 PITKÄNEN, TUULLI, Alcohol drinking behavior and its developmental antecedents. - Alkoholien juomiskäyttäytyminen ja sen ennustaminen. 103 p. (169 p.) Tiivistelmä 6 p. 2006.
- 294 LINNILÄ, MAIJA-LIISA, Kouluvalmiudesta koulun valmiuteen. Poikkeuksellinen koulunaloitus koulumenestyksen, viranomaislausuntojen ja perheiden kokemusten valossa. - From school readiness to readiness of school - Exceptional school starting in the light of school attainment, official report and family experience. 321 p. Summary 3 p. 2006.
- 295 LEINONEN, ANU, Vanhusneuvoston funktioita jäljittämässä. Tutkimus maaseutumaisten kuntien vanhusneuvostoista. - Tracing functions of older people's councils. A study on older people's councils in rural municipalities. 245 p. Summary 3 p. 2006.
- 296 KAUPPINEN, MARKO, Canon vs. charisma. "Maoism" as an ideological construction. - Kaanon vs. karisma. "Maoismi" ideologisenä konstruktiona. 119 p. Yhteenveto 2 p. 2006.
- 297 VEHKAKOSKI, TANJA, Leimattu lapsuus? Vammaisuuden rakentuminen ammatti-ihmisten puheessa ja teksteissä. - Stigmatized childhood? Constructing disability in professional talk and texts. 83 p. (185 p.) Summary 4 p. 2006.
- 298 LEPPÄÄHO, HENRY, Matemaattisen ongelman ratkaisutaidon opettaminen peruskoulussa. Ongelmanratkaisukurssin kehittäminen ja arviointi. - Teaching mathematical problem solving skill in the Finnish comprehensive school. Designing and assessment of a problem solving course. 343 p. Summary 4 p. 2007.
- 299 KUVAJA, KRISTIIINA, Living the Urban Challenge. Sustainable development and social sustainability in two southern megacities. 130 p. (241 p.) Yhteenveto 4 p. 2007.
- 300 POIJOLA, PASI, Technical artefacts. An ontological investigation of technology. 150 p. Yhteenveto 3 p. 2007.
- 301 KAUKUA, JARI, Avicenna on subjectivity. A philosophical study. 161 p. Yhteenveto 3 p. 2007.
- 302 KUPILA, PÄIVI, "Minäkö asiantuntija?". Varhaiskasvatuksen asiantuntijan merkitysperspektiivin ja identiteetin rakentuminen. - "Me, an expert?" Constructing the meaning perspective and identity of an expert in the field of early childhood education. 190 p. Summary 4 p. 2007.
- 303 SILVENNOINEN, PIIA, Ikä, identiteetti ja ohjaava koulutus. Ikääntyvät pitkäaikaistyöttömät oppimisyhteiskunnan haasteena. - Age, identity and career counselling. The ageing, long-term unemployed as a challenge to learning society. 229 p. Summary 4 p. 2007.
- 304 REINIKAINEN, MARJO-RIITTA, Vammaisuuden sukupuolittuneet ja sortavat diskurssit: Yhteiskunnallis-diskursiivinen näkökulma vammaisuuteen. - Gendered and oppressive discourses of disability: Social-discursive perspective on disability. 81 p. (148 p.) Summary 4 p. 2007.
- 305 MÄÄTTÄ, JUKKA, Asepalvelus nuorten naisten ja miesten opinto- ja työuralla. - The impact of military service on the career and study paths of young women and men. 141 p. Summary 4 p. 2007.
- 306 PYYKKÖNEN, MIIKKA, Järjestäytyvät diasporat. Etnisyys, kansalaisuus, integraatio ja hallinta maahanmuuttajien yhdistystoiminnassa. - Organizing diasporas. Ethnicity, citizenship, integration, and government in immigrant associations. 140 p. (279 p.) Summary 2 p. 2007.
- 307 RASKU, MINNA, On the border of east and west. Greek geopolitical narratives. - Idän ja lännen rajalla. Narratiiveja kreikkalaisesta geopolitiikasta. 169 p. Yhteenveto 3 p. 2007.
- 308 LAPIOLAHTI, RAIMO, Koulutuksen arviointi kunnallisen koulutuksen järjestäjän tehtävänä. Paikallisen arvioinnin toteutumisedellytysten arviointia erään kuntaorganisaation näkökulmasta. - The evaluation of schooling as a task of the communal maintainer of schooling - what are the presuppositions of the execution of evaluation in one specific communal organization. 190 p. Summary 7 p. 2007.
- 309 NATALE, KATJA, Parents' Causal Attributions Concerning Their Children's Academic Achievement. - Vanhempien lastensa koulu-menestystä koskevat kausaaliattribuutit. 54 p. (154 p.) Yhteenveto 3 p. 2007.
- 310 VAHTERA, SIRPA, Optimistit opintiellä. Opinnoissaan menestyvien nuorten hyvinvointi lukiosta jatko-opintoihin. - The well-being of optimistic, well-performing high school students from high school to university. 111 p. Summary 2 p. 2007.
- 311 KOIVISTO, PÄIVI, "Yksilöllistä huomiota arkisissa tilanteissa". Päiväkodin toimintakulttuurin kehittäminen lasten itsetuntoa vahvistavaksi. - "Individual attention in everyday situations". Developing the operational culture of a day-care centre to strengthen children's self-esteem. 202 p. Summary 4 p. 2007.
- 312 LAHIKAINEN, JOHANNA, "You look delicious" - Food, eating, and hunger in Margaret Atwood's novels. 277 p. Yhteenveto 2 p. 2007.
- 313 LINNAVUORI, HANNARIKKA, Lasten kokemuksia vuoroasumisesta. - Children's experiences of dual residence. 202 p. Summary 8 p. 2007.
- 314 PARVIAINEN, TIINA, Cortical correlates of language perception. Neuromagnetic studies in adults and children. - Kielen käsittely aivoissa. Neuromagneettisia tutkimuksia aikuisilla ja lapsilla. 128 p. (206 p.) Yhteenveto 5 p. 2007.



- 315 KARA, HANNELE, *Ermutige mich Deutsch zu sprechen. Portfolio als evaluationsform von mündlichen leistungen.* - "Rohkaise minua puhumaan saksaa" - kielisalkku suullisen kielitaidon arviointivälineenä. 108 p. Yhteenveto 3 p. 2007.
- 316 MÄKELÄ, AARNE, *Mitä rehtorit todella tekevät. Etnografinen tapaustutkimus johtamisesta ja rehtorin tehtävistä peruskoulussa.* - What principals really do. An ethnographic case study on leadership and on principal's tasks in comprehensive school. 266 p. Summary 5 p. 2007.
- 317 PUOLAKANAHO, ANNE, *Early prediction of reading - Phonological awareness and related language and cognitive skills in children with a familial risk for dyslexia.* - Lukemistaitojen varhainen ennustaminen. Fonologinen tietoisuus, kielelliset ja kognitiiviset taidot lapsilla joiden suvussa esiintyy dysleksiaa. 61 p. (155 p.) Yhteenveto 3 p. 2007.
- 318 HOFFMAN, DAVID M., *The career potential of migrant scholars in Finnish higher education. Emerging perspectives and dynamics.* - Akateemisten siirtolaisten uramahdollisuudet suomalaisessa korkeakoulujärjestelmässä: dynamiikkaa ja uusia näkökulmia. 153 p. (282 p.) Yhteenveto 2 p. 2007.
- 319 FADJUKOFF, PÄIVI, *Identity formation in adulthood.* - Identiteetin muotoutuminen aikuisiässä. 71 p. (168 p.) Yhteenveto 5 p. 2007.
- 320 MÄKIKANGAS, ANNE, *Personality, well-being and job resources: From negative paradigm towards positive psychology.* - Persoonallisuus, hyvinvointi ja työn voimavarat: Kohti positiivista psykologiaa. 66 p. (148 p.) Yhteenveto 3 p. 2007.
- 321 JOKISAARI, MARKKU, *Attainment and reflection: The role of social capital and regrets in developmental regulation.* - Sosiaalisen pääoman ja toteutumattomien tavoitteiden merkitys kehityksen säätelyssä. 61 p. (102 p.) Yhteenveto 2 p. 2007.
- 322 HÄMÄLÄINEN, JARMO, *Processing of sound rise time in children and adults with and without reading problems.* - Äänten nousuaikojen prosessointi lapsilla ja aikuisilla, joilla on dysleksia ja lapsilla ja aikuisilla, joilla ei ole dysleksiaa. 48 p. (95 p.) Tiivistelmä 2 p. 2007.
- 323 KANERVIO, PEKKA, *Crisis and renewal in one Finnish private school.* - Kriisi ja uudistuminen yhdessä suomalaisessa yksityiskoulussa. 217 p. Tiivistelmä 2 p. 2007.
- 324 MÄÄTTÄ, SAMI, *Achievement strategies in adolescence and young adulthood.* - Nuorten ajattelu- ja toimintastrategia. 45 p. (120 p.) Tiivistelmä 3 p. 2007.
- 325 TORPPA MINNA, *Pathways to reading acquisition: Effects of early skills, learning environment and familial risk for dyslexia.* - Yksilöllisiä kehityspolkuja kohti lukemisen taitoa: Varhaisten taitojen, oppimisympäristön ja sukuriskin vaikutukset. 53 p. (135 p.) 2007.
- 326 KANKAINEN, TOMI, *Yhdistykset, instituutiot ja luottamus.* - Voluntary associations, institutions and trust. 158 p. Summary 7 p. 2007.
- 327 PIRNES, ESA, *Merkityksellinen kulttuuri ja kulttuuripolitiikka. Laaja kulttuurin käsite kulttuuripolitiikan perusteluna.* - Meaningful culture and cultural policy. A broad concept of culture as a basis for cultural policy. 294 p. Summary 2 p. 2008.
- 328 NIEMI, PETTERI, *Mieli, maailma ja referenssi. John McDowellin mielenfilosofian ja semantiikan kriittinen tarkastelu ja ontologinen täydennys.* - Mind, world and reference: A critical examination and ontological supplement of John McDowell's philosophy of mind and semantics. 283 p. Summary 4 p. 2008.
- 329 GRANBOM-HERRANEN, LIISA, *Sananlaskut kasvatustieteessä - perinnettä, kasvatusta, indoktrinaatiota? - Proverbs in pedagogical discourse - tradition, upbringing, indoctrination?* 324 p. Summary 8 p. 2008.
- 330 KYKYRI, VIRPI-LIISA, *Helping clients to help themselves. A discursive perspective to process consulting practices in multi-party settings.* - Autetaan asiakasta auttamaan itse itseään. Diskursiivinen näkökulma prosessikonsultoinnin käytäntöihin ryhmätilanteissa. 75 p. (153 p.) Tiivistelmä 2 p. 2008.
- 331 KIURU, NOONA, *The role of adolescents' peergroups in the school context.* - Nuortentoveriryhmien rooli kouluympäristössä. 77 p. (192 p.) Tiivistelmä 3 p. 2008.
- 332 PARTANEN, TERHI, *Interaction and therapeutic interventions in treatment groups for intimately violent men.* 46 p. (104 p.) Yhteenveto 2 p. 2008.
- 333 RAITILA, RAIJA, *Retkellä. Lasten ja kaupunkiympäristön kohtaaminen.* - Making a visit. Encounters between children and an urban environment. 179 p. Summary 3 p. 2008.
- 334 SUME, HELENA, *Perheen pyörteinen arki. Sisäkorvaistutetta käyttävän lapsen matka kouluun.* - Turbulent life of the family. Way to school of a child with cochlear implant. 208 p. Summary 6 p. 2008.
- 335 KOTIRANTA, TUUJA, *Aktivoinnin paradoksit.* - The paradoxes of activation. 217 p. Summary 3 p. 2008.
- 336 RUOPPILA, ISTO, HUUHTANEN, PEKKA, SEITSAMO, JORMA AND ILMARINEN, JUHANI, *Age-related changes of the work ability construct and its relation to cognitive functioning in the older worker: A 16-year follow-up study.* 97 p. 2008.
- 337 TIKKANEN, PIRJO, *"Helpompaa ja hausempaa kuin luulin". Matematiikka suomalaisten ja unkarilaisten perusopetuksen neljäsluokkalaisten kokemana.* - "Easier and more fun that

- I thought". Mathematics experienced by fourth-graders in Finnish and Hungarian comprehensive schools. 309 p. Summary 3 p. 2008.
- 338 KAUPPINEN, ILKKA, Tiedon omistaminen on valtaa – Globalisoituvan patenttijärjestelmän poliittinen moraalilaitos ja globaali kapitalismi. – *Owning* knowledge is power. Political moral economy of the globalizing patent system and global capitalism. 269 p. Summary 5 p. 2008.
- 339 KUJALA, MARIA, Muukalaisena omassa maassa. Miten kasvaa vuorovaikutuskonflikteissa? – A stranger in one's own land. How to grow in interaction conflicts? 174 p. Summary 7 p. 2008.
- 340 KOPONEN, TUIRE, Calculation and Language: Diagnostic and intervention studies. – Laskutaito ja kieli: Diagnostinen ja kuntoutustutkimus. 49 p. (120 p.) Tiivistelmä 2 p. 2008.
- 341 HAUTALA, PÄIVI-MARIA, Lupa tulla näkyväksi. Kuvataideterapeuttinen toiminta kouluissa. – Permission to be seen. Art therapeutic activities in schools. 202 p. 2008.
- 342 SIPARI, SALLA, Kuntouttava arki lapsen tueksi. Kasvatuksen ja kuntoutuksen yhteistoiminnan rakentuminen asiantuntijoiden keskusteluissa. – Habilitative everyday life to support the child. Construction of the collaboration of education and rehabilitation in experts discussions. 177 p. Summary 4 p. 2008.
- 343 LEHTONEN, PÄIVI HANNELE, Voimauttava video. Asiakslähtöisyyden, myönteisyyden ja videokuvan muodostama työorientaatio perhetyön menetelmänä. – Empowering video. A work orientation formed by client-focus, positivity and video image as a method for family work. 257 p. Summary 3 p. 2008.
- 344 RUOHOMÄKI, JYRKI, "Could Do Better". Academic Interventions in Northern Ireland Unionism. – "Could Do Better" Akateemiset interventiot Pohjois-Irlannin unionismiin. 238 p. Tiivistelmä 2 p. 2008.
- 345 SALMI, PAULA, Nimeäminen ja lukemisvaikeus. Kehityksen ja kuntoutuksen näkökulma. – Naming and dyslexia: Developmental and training perspectives. 169 p. Summary 2 p. 2008.
- 346 RANTANEN, JOHANNA, Work-family interface and psychological well-being: A personality and longitudinal perspective. – Työn ja perheen vuorovaikutuksen yhteys psyykkiseen hyvinvointiin sekä persoonallisuuteen pitkittäistutkimuksen näkökulmasta 86 p. (146 p.) Yhteenveto 6 p. 2008.
- 347 PIIPPO, JUKKA, Trust, Autonomy and Safety at Integrated Network- and Family-oriented mode for co-operation. A Qualitative Study. 70 p. (100 p.) Yhteenveto 2 p. 2008.
- 348 HÄTINEN, MARJA, Treating job burnout in employee rehabilitation: Changes in symptoms, antecedents, and consequences. – Työuupumuksen hoito työikäisten kuntoutuksessa: muutokset työuupumuksen oireissa, ennakoijissa ja seurauksissa. 85 p. (152 p.) Tiivistelmä 4 p. 2008.
- 349 PRICE, GAVIN, Numerical magnitude representation in developmental dyscalculia: Behavioural and brain imaging studies. 139 p. 2008.
- 350 RAUTAINEN, MATTI, Keiden koulu? Aineenopettajaksi opiskelevien käsityksiä koulu-kulttuuriin yhteisöllisyydestä. – Who does school belong to? Subject teacher students' conceptions of community in school culture. 180 p. Summary 4 p. 2008.
- 351 UOTINEN, SANNA, Vanhempien ja lasten toimijuuteen konduktiivisessa kasvatuksessa. – Into the agency of a parent and a child in conductive education. 192 p. Summary 3 p. 2008.
- 352 AHONEN, HELENA, Rehtoreiden kertoma johtajuus ja johtajaidentiteetti. – Leadership and leader identity as narrated by headmasters. 193 p. 2008.
- 353 MOISIO, OLLI-PEKKA, Essays on radical educational philosophy. 151 p. Tiivistelmä 3 p. 2009.
- 354 LINDQVIST, RAIIJA, Parisuhdeväkivallan kohtaaminen maaseudun sosiaalityössä. – Encountering partner violence with rural social work. 256 p. 2009.
- 355 TAMMELIN, MIA, Working time and family time. Experiences of the work and family interface among dual-earning couples in Finland. – Työaika ja perheen aika: kokemuksia työn ja perheen yhteensovittamisesta Suomessa. 159 p. Tiivistelmä 3 p. 2009.
- 356 RINNE, PÄIVI, Matkalla muutokseen. Sosiaalialan projektitoiminnan perustelut, tavoitteet ja toimintatavat Sosiaaliturva-lehden kirjoituksissa 1990-luvulla. – On the way to the change. 221 p. Summary 2 p. 2009.
- 357 VALTONEN, RIITTA, Kehityksen ja oppimisen ongelmien varhainen tunnistaminen Lene-arvioinnin avulla. Kehityksen ongelmien päällekkäisyys ja jatkuvuus 4–6-vuotiailla sekä ongelmien yhteys koulusuoriutumiseen. – Lene-assessment and early identification of developmental and learning problems. Co-occurrence and continuity of developmental problems from age 4 to age 6 and relation to school performance. 73 p. (107 p.) Summary 2 p. 2009.
- 358 SUHONEN, KATRI, Mitä hiljainen tieto on hengellisessä työssä? Kokemuksellinen näkökulma hiljaisen tiedon ilmenemiseen, siirrettävyyteen ja siirrettävyyden merkitykseen ikääntyneiden diakoniatyöntekijöiden ja pappien työssä. – What is tacit knowledge in spiritual work? An experiential approach to the manifestation, significance and distribution of tacit knowledge in the work of aged church deacons and ministers. 181 p. Summary 6 p. 2009.

- 359 JUMPPANEN, AAPO, United with the United States - George Bush's foreign policy towards Europe 1989-1993. 177 p. Yhteenveto 3 p. 2009.
- 360 HUEMER, SINI, Training reading skills. Towards fluency. - Lukemistaitojen harjoittaminen. Tavoitteena sujuvuus. 85 p. (188 p.) Yhteenveto 3 p. 2009.
- 361 ESKELINEN, TEPPO, Putting global poverty in context. A philosophical essay on power, justice and economy. 221 p. Yhteenveto 1 p. 2009.
- 362 TAIPALE, SAKARI, Transformative technologies, spatial changes: Essays on mobile phones and the internet. 97 p. (184 p.) Yhteenveto 3 p. 2009.
- 363 KORKALAINEN, PAULA, Riittämättömyyden tunteesta osaamisen oivallukseen. Ammatillisen asiantuntijuuden kehittäminen varhaiserityiskasvatuksen toimintaympäristöissä. - From a feeling of insufficiency to a new sense of expertise. Developing professional knowledge and skills in the operational environments for special needs childhood education and care. 303 p. Summary 4 p. 2009.
- 364 SEPPÄLÄ-PÄNKÄLÄINEN, TARJA, Oppijoiden moninaisuuden kohtaaminen suomalaisessa lähikoulussa. Etnografia kouluyhteisön aikuisten yhdessä oppimisen haasteista ja mahdollisuuksista. - Confronting the Diversity of Learners in a Finnish Neighbourhood School. An Ethnographic Study of the Challenges and Opportunities of Adults Learning Together in a School community. 256 p. Summary 4 p. 2009.
- 365 SEVÓN, EIJA, Maternal Responsibility and Changing Relationality at the Beginning of Motherhood. - Äidin vastuu ja muuttuvat perhesuhteet äitiyden alussa. 117 p. (200 p.) Yhteenveto 5 p. 2009.
- 366 HUUTTUNEN-SCOTT, TIINA, Auditory duration discrimination in children with reading disorder, attention deficit or both. - Kuulonvarainen keston erottelu lapsilla, joilla on lukemisvaikeus, tarkkaavaisuuden ongelma tai molemmat. 68 p. (112 p.) Tiivistelmä 3 p. 2009.
- 367 NEUVONEN-RAUHALA, MARJA-LIISA, Työelämälähtöisyyden määrittäminen ja käyttäminen ammattikorkeakoulun jatkotutkintokokeilussa. - Defining and applying working-life orientation in the polytechnic postgraduate experiment. 163 p. Summary 7 p. 2009.
- 368 NYMAN, TARJA, Nuoren vieraan kielen opettajan pedagogisen ajattelun ja ammatillisen asiantuntijuuden kehittyminen. - The development of pedagogical thinking and professional expertise of newly qualified language teachers. 121 p. (201 p.) Summary 4 p. 2009.
- 369 PUUTIO, RISTO, Hidden agendas. Situational tasks, discursive strategies and institutional practices in process consultation. 83 p. (147 p.) Tiivistelmä 2 p. 2009.
- 370 TOIVANEN, JUHANA, Animal consciousness. Peter Olivi on cognitive functions of the sensitive soul. 369 p. Yhteenveto 4 p. 2009.
- 371 NOKIA, MIRIAM, The role of the hippocampal theta activity in classical eyeblink conditioning in rabbits. - Hippokampuksen theta-aktiivisuuden rooli klassisessa silmäniskuehdollistamisessa kaneilla. 41 p. (80 p.) Yhteenveto 2 p. 2009.
- 372 LÄHTEENMÄKI, VILI, Essays on early modern conceptions of consciousness: Descartes, Cudworth, and Locke. 160 p. 2009.
- 373 BJÖRK, KAJ, What explains development. Development strategy for low human development index countries. 212 p. Yhteenveto 1 p. 2009.
- 374 PUUPPONEN, ANTTI, Maaseutuyrittäjyys, verkostot ja paikallisuus. Tapaustutkimus pienimuotoisen elintarviketuotannon kestävyyydestä Keski-Suomessa. - Rural entrepreneurship, networks and locality. A case study of the sustainability of small-scale food production in Central Finland. 100 p. (191 p.) Summary 3 p. 2009.
- 375 HALTUNEN, LEENA, Päivähoitotyö ja johtajuus hajautetussa organisaatiossa. - Day care work and leadership in a distributed organization. 181 p. Summary 4 p. 2009.
- 376 KAIDESOJA, TUUKKA, Studies on ontological and methodological foundations of critical realism in the social sciences. 65 p. (187 p.) Yhteenveto 9 p. 2009.
- 377 SIPPOLA, MARKKU, A low road to investment and labour management? The labour process at Nordic subsidiaries in the Baltic States. 272 p. Tiivistelmä 2 p. 2009.
- 378 SANTALA, OLLI-PEKKA, Expertise in using the Rorschach comprehensive system in personality assessment. 150 p. Tiivistelmä 1 p. 2009.
- 379 HARJUNEN, HANNELE, Women and fat: Approaches to the social study of fatness. - Naiset ja lihavuus: näkökulmia lihavuuden yhteiskuntatieteelliseen tutkimukseen 87 p. (419 p.) Tiivistelmä 4 p. 2009.
- 380 KETTUNEN, LIISA, Kyllä vai ei. Peruskoulun sukupuolikasvatuksen oppimateriaalin kehittämistyö ja arviointi. - Yes or no? The development and evaluation of teaching material for sex education in the Finnish comprehensive school. 266 p. Summary 3 p. 2010.
- 381 FROM, KRISTINE, "Että sais olla lapsena toisten lasten joukossa". Substantiivinen teoria erityistä tukea tarvitsevan lapsen toiminnallisesta osallistumisesta toimintaympäristöissä. - To be a child just as the others in the peer group. A substantive theory of activity-

- based participation of the child with special educational needs. 174 p. Summary 4 p. 2010.
- 382 MYKKÄNEN, JOHANNA, Isäksi tulon tarinat, tunteet ja toimijuus. - Becoming a father – types of narrative, emotions and agency. 166 p. Summary 5 p. 2010.
- 383 RAASUMAA, VESA, Perusopetuksen rehtori opettajien osaamisen johtajana. - Knowledge management functions of a principal in basic education. 349 p. Summary 5 p. 2010.
- 384 SISÄINEN, LAURI, Foucault's voices: Toward the political genealogy of the auditory-sonorous. - Foucault'n äänet. Kohti auditoris-sonoorista poliittista genealogiaa. 207 p. Tiivistelmä 2 p. 2010.
- 385 PULLI, TUULA, Totta ja unta. Draama puhe- ja kehitysvammaisten ihmisten yhteisöllisenä kuntoutuksena ja kokemuksena. - The Real and the Illusory. Drama as a means of community-based rehabilitation and experience for persons with severe learning and speech disabilities. 281 p. Summary 7 p. 2010.
- 386 SISKONEN, TIINA, Kielelliset erityisvaikeudet ja lukemaan oppiminen. - Specific language impairments and learning to read. 205 p. Summary 3 p. 2010.
- 387 LYYRA, PESSI, Higher-order theories of consciousness: An appraisal and application. - Korkeamman kertaluvun tietoisusteoria: arvio ja käyttöehdotus. 163 p. Yhteenveto 5 p. 2010.
- 388 KARJALAINEN, MERJA, Ammattilaisten käsityksiä mentoroinnista työpaikalla. - Professionals' conceptions of mentoring at work. 175 p. Summary 7 p. 2010.
- 389 GEMECHU, DEREJE TEREFE, The implementation of a multilingual education policy in Ethiopia: The case of Afaan Oromoo in primary schools of Oromia Regional State. 266 p. 2010.
- 390 KOIVULA, MERJA, Lasten yhteisöllisyys ja yhteisöllinen oppiminen päiväkodissa. - Children's sense of community and collaborative learning in a day care centre. 189 p. Summary 3 p. 2010.
- 391 NIEMI, MINNA, Moraalijärjestystä tuottamassa. Tutkimus poliisityöstä lasten parissa. - Producing moral order. A Study on police work with children. 190 p. Summary 3 p. 2010.
- 392 ALEMAYEHU TEKLEMARIAM HAYE, Effects of intervention on psychosocial functioning of hearing and hard of hearing children in selected primary schools of Addis Ababa, Ethiopia. 195 p. Executive summary 4 p. 2010.
- 393 KASKIHARJU, EIJA, Koteja ja kodinomaisuutta. Tutkimus vanhenemisen paikoista valtiopäiväpuheissa 1950 - 2005. - Homes and homelikeness. A study on places for ageing in parliamentary speeches from 1950 to 2005. 244 p. Summary 5 p. 2010.
- 394 MAHLAKAARTO, SALME, Subjektiksi työssä - Identiteettiä rakentamassa voimaantumisen kehitysohjelmassa. - Becoming a subject at work - Constructing identity within a program of empowerment. 95 p. (198 p.) Yhteenveto 1 p. 2010.
- 395 TAPIO, TARJA, "Meilä on kaikilla samanlaiset tarinat". Tarinankerrontatutkimus tornionlaaksoisuudesta vanhimpien aapulaisten arjessa ja tulevaisuudessa. - "We all have the same stories". A storytelling case study of Torne Valley -ness in the everyday life and future of elderly Aapua residents. 261 p. Summary 6 p. 2010.
- 396 RAUTAINEN, EIJA-LIISA, Co-construction and collaboration in couple therapy for depression. - Yhteistoiminnallisuus masennuksen pariterapiassa. 56 p. (122 p.) Yhteenveto 3 p. 2010.
- 397 AALTONEN, TERHI, "Taiteilija ei vanhene". Haastattelututkimus kuvataiteilijoiden ikääntymiskokemuksista taidemaailmassa. - "An artist doesn't get old". An interview-based study of painters' experiences of ageing in the world. 216 p. Summary 5 p. 2010.
- 398 SAVOLAINEN, KAISA, Education as a means to world peace: The case of the 1974 UNESCO recommendation. - Kasvatus maailmanrauhan välineenä: Tapaustutkimus UNESCON 1974 hyväksymästä suosituksesta. 262 p. Yhteenveto 12 p. 2010.
- 399 HEMMINKI, ARJA, Kertomuksia avioerosta ja parisuhteen päättämisestä. Suomalainen eropuhe pohjalaisten kirjoituksissa ja naistenlehdissä. - Narratives on divorce and ending of a relationship. 158 p. Summary 2 p. 2010.
- 400 SAINI, NINA, On the rocky road of reading: Effects of computer-assisted intervention for at-risk children. - Lukemaan oppimisen kivisellä tiellä - Verkkopohjaisen Ekapeli - ohjelman kuntouttavat vaikutukset riskilasten lukemaan oppimisessä. 95 p. (208 p.) Yhteenveto 5 p. 2010.
- 401 VILJARANTA, JAANA, The development and role of task motivation and task values during different phases of the school career. - Oppiainekohtaisen koulumotivaation kehitys ja rooli koulutaipaleen eri vaiheissa. 53 p. (115 p.) Yhteenveto 1 p. 2010.
- 402 OINAS, TOMI, Sukupuolten välinen kotityönjako kahden ansaitsijan perheissä. - Domestic division of labour in dual-earner households. 188 p. 2010.
- 403 MAMMON, REET, Kolmen etnisen ryhmän kotoutumisprosessi Suomessa. - The integration process of three ethnic groups in Finland. 142 p. Summary 5 p. 2010.
- 404 KETONEN, RITVA, Dysleksiariski oppimisen haasteena. Fonologisen tietoisuuden interventio ja lukemaan oppiminen. - Risk for dyslexia as a challenge of learning. Phonological intervention and learning to read. 139 p. Summary 3 p. 2010.

- 405 LAHTERO, TAPIO, Yhtenäiskoulun johtamiskulttuuri. Symbolis-tulkinnallinen näkökulma. - Leadership culture in unified comprehensive school, symbolic-interpretative approach. 238 p. Summary 2 p. 2011.
- 406 NÖTKÖ, MARIANNE, Väkiältä, vallankäyttö ja vahingoittuminen naisten perhesuhteissa. - Violence, power using and being hurt in women's family relations. 254 p. Summary 5 p. 2011.
- 407 PULKKINEN, SEPPÖ, Valmentajataustan merkitys rehtorin työssä. - The significance of coaching background in principal's work. 211 p. Summary 7 p. 2011.
- 408 SÖDÖR, UUVE, Community resilience and wellbeing in northwest Russian forestry settlements. 195 p. Summary 2 p. 2011.
- 409 HYVÖNEN, KATRIINA, Personal work goals put into context: Associations with work environment and occupational well-being. - Henkilökohtaisten työtavoitteiden puitteet: yhteydet työoloihin ja työhyvinvointiin 82 p. (133 p.) Yhteenveto 3 p. 2011.
- 410 RUOHOTIE-LYHTY, MARIA, Opettajuuden alkutaival. Vastavalmistuneen vieraan kielen opettajan toimijuus ja ammatillinen kehittyminen. - First steps on the path of teacherhood. Newly qualified foreign language teachers' agency and professional development. 98 p. (190 p.) Summary 2 p. 2011.
- 411 HALONEN, KATRI, Kulttuurituottajat taiteen ja talouden risteyskohdassa. - Cultural intermediaries at the junction between art and business. 101 p. (175 p.) Summary 6 p. 2011.
- 412 MIKOLA, MARJATTA, Pedagogista rajankäyntiä koulussa. Inklusioreitit ja yhdessä oppimisen edellytykset. - Defining pedagogical boundaries at school - the routes to inclusion and conditions for collaborative learning. 304 p. Summary 8 p. 2011.
- 413 SOANJÄRVI, KATARINA, Mitä on ammatillinen nuorisotyö? Nuorisotyön villiä kenttää kesyttämässä. - What is professional youth work? Taming of wild youth work sector. 152 p. Summary 3 p. 2011.
- 414 LEHTO-SALO, PIRKKO, Koulukotisijoitus - nuoren toinen mahdollisuus? Mielenterveyden häiriöiden, oppimisvaikeuksien ja perheongelmien kirjo kehittämishaasteena. - The placement on a reform school - The second change for an adolescent? Psychiatric disorders, learning disabilities and problems in the family as a challenge of treatment. 74 p. (125 p.) 2011.
- 415 MÄGI, KATRIN, Achievement goals, achievement behaviours, and skill development in the school and home context: Their antecedents and correlates. 54 p. (134 p.) Yhteenveto 3 p. 2011.
- 416 HEINONEN, VEIKKO U.J., Suomalaisen turvallisuuspolitiikan tila. Suomalaisen 2000-luvun alun turvallisuuspoliittisen debatin käsiteanalyttinen tarkastelu. - The state of Finnish security policy. A conceptual analysis of the Finnish debate on security policy in the early 2000s. 266 p. Yhteenveto 1 p. 2011.
- 417 PURO, ERIKA, Peruskoulun erityisopetuksen laatu oppilaiden, huoltajien, koulunkäyntiavustajien, opettajien ja rehtorien arvioimana. - The quality of special education in Finnish comprehensive schools from the viewpoint of pupils, parents, educational assistants, teachers and principals. 165 p. Summary 2 p. 2011.
- 418 PIETIKÄINEN, MAIJA, Laulu ja amor mundi. Tutkimus laulusta eräänä maailmasta huolettimisen perusaktiviteettina. - Singing and amor mundi. 193 p. Summary 2 p. 2011.
- 419 SALO, SIRPA, Parisuhdeongelmat ja lasten psyykinen hyvinvointi: kaksi tutkimuskäsitelmää. - Couples conflicts and children's mental wellbeing: two research approaches. 156 p. Summary 2 p. 2011.
- 420 HANNONEN, RIITTA, Verbal and academic skills in children with type 1 diabetes. - Tyypin 1 diabetesta sairastavien lasten kielelliset ja oppimiseen liittyvät perustaidot. 49 p. (79 p.) Yhteenveto 3 p. 2011.
- 421 SELIN, JANI, Hallinnan näkökulmia huumeriippuvuuden hoitoon Suomessa vuosina 1965-2005. - Governmental perspectives on the treatment of drug addiction in Finland between 1965 and 2005. 55 p. (118 p.) Summary 2 p. 2011.
- 422 NIKKOLA, TIINA, Oppimisen esteet ja mahdollisuudet ryhmässä. Syyllisyyden kehittyminen syntipukki-ilmiöksi opiskeluryhmässä ohjaajan tulkitsemana. - Learning in a group: obstacles and opportunities. A supervisor's interpretation of how guilt becomes scapegoating in a study group. 244 p. Summary 6 p. 2011.
- 423 LASSENIUS, YVONNE, Utvecklingen av vårdlärarnas kulturkompetens. En kvalitativ studie som fokuserar på lärarna i mångkulturella lärandemiljöer i Finland. - The development of nurse teachers' cultural competence. A qualitative study focusing on teachers in multicultural learning environments in Finland. 74 p. (101 p.) Summary 2 p. 2011.
- 424 YLÖNEN, MARJA, Saastumisen kontrollin ideologia. Vesien saastumisen ja vesirikosten kontrolli Suomessa vuosina 1960-2000. - Ideology of pollution control. Social control of water pollution and pollution crimes in Finland from the 1960s until the year 2000. 232 p. Summary 2 p. 2011.
- 425 PYLKKÄ, OUTI, Duuri- ja mollisointuja musiikkialan kentällä. Organisaatiomuutoksen heijastuksia ammattikorkeakoulun musiikinopettajien työhön. - Ideology of pollution control. Major and minor

- chords in the field of music profession reflections of organizational changes in the music teachers' work at the University of Applied Sciences. 190 p. Summary 4 p. 2011.
- 426 MALINEN, KAISA, Parisuhde pikkulapsiperheessä. - Spousal relationship in families with young children. 79 p. (147 p.) Summary 4 p. 2011.
- 427 MURTORINNE-LAHTINEN, MINNA, Äitipuolen identiteetin rakentaminen uusperheessä. - Identity construction of stepmothers in a stepfamily. 222 p. Summary 6 p. 2011.
- 428 RUOKOLAINEN, MERVI, Do organizational and job-related factors relate to organizational commitment? A mixed method study of the associations. - Ovatko organisaation ja työn piirteet yhteydessä organisaatioon sitoutumiseen? Monimenetelmällinen tutkimus ilmiöiden välisistä yhteyksistä. 135 p. Yhteenveto 3 p. 2011.
- 429 JÄRVILEHTO, LAURI, Pragmatic a priori knowledge: A pragmatic approach to the nature and object of what can be known independently of experience. 153 p. 2011.
- 430 RUOPPILA, ISTO & IIVANAINEN, MATTI (EDS.), Life Course of Finnish Persons with Intellectual Disabilities Born 1898-1960. A 35-Year Follow-up Study. 119 p. Yhteenveto 9 p. 2011.
- 431 BJÖRK, ANNA, The politics of citizenship tests: time, integration and the contingent polity. 123 p. Yhteenveto 1 p. 2011.
- 432 SAARINEN, MINNA, Sosiaalisen elämän kehukset. Kampus-ohjelman opiskelijoiden sosiaalinen asema tuetussa aikuisopiskelussa ja vapaa-ajalla. - Framework of social life. The social life of Campus-program students in their supported adult learning and leisure. 276 p. 2012.
- 433 HÖKKÄ, PÄIVI, Teacher educators amid conflicting demands: tensions between individual and organizational development. - Opettajankouluttajat ristiriitaisten vaatien keskellä: yksilön ja organisaation välisiä jännitteitä. 127p. (197 p.) 2012.
- 434 SALMINEN, VELI-MATTI, Verkostot, paikallisuus ja eriarvoisuus. - Networks, locality and inequality. 160 p. Summary 6 p. 2012.
- 435 HAUTALA, JARKKO, Visual word recognition in fluent and dysfluent readers in the transparent Finnish orthography. - Visuaalinen sanantunnistus sujuvilla ja hitailla lukijoilla kirjain-äänne vastaavuudeltaan läpinäkyvässä suomen kielessä. 54 p. (132 p.) Tiivistelmä 3 p. 2012.
- 436 SILINSKAS, GINTAUTAS, Parental involvement and children's academic skills. 70 p. (153 p.) Tiivistelmä 1 p. 2012.
- 437 RONIMUS, MIIA, Digitaalisen oppimispelin motivoivuus. Havaintoja Ekapeliä pelanneista lapsista. - The motivational appeal of a digital learning game. Observations of children using Graphogame. 163 p. Summary 5 p. 2012.
- 438 KAUTTO-KNAPE, ERJA, Oppilasta lamaannuttava kouluvuorovaikutus: aineistoperustainen teoria. - The paralysing school interaction: A grounded theory study of the students' view. 151 p. Summary 2 p. 2012.
- 439 KNIF, PIRJO, Keski-ikäiset ikääntyvien omaistensa auttajina? Tutkimus sukupolvien välisestä informaalista hoivasta. - The middle-aged as caregivers for the elderly relatives? A study of informal care between generations. 200 p. Summary 3 p. Sammanfattning 2 p. 2012.
- 440 KULOVAARA, KARI, Euroopan ammatillinen yhteisjärjestö EAY ja strategiset valinnat: identiteetit, ideologiat ja diskurssit. - European Trade Union Confederation ETUC and strategic choices. Identities, ideologies and discourses. 249 p. Summary 2 p. 2012.
- 441 KIILLI, CARITA, Online reading as an individual and social practice. 72 p. (163 p.) Summary 4 p. 2012.
- 442 SAMA, THOMAS BABILA, NGOs and the implementation of active labour market policies: the case of Finland and Sweden. 104 p. (206 p.) Yhteenveto 2 p. 2012.
- 443 OKSANEN, ANNAMAIJA, Auditiivisen ja kielellisen harjoittelun vaikutus kolmannella luokalla olevien heikkojen lukijoiden lukutaitoon. - Speech and non-speech training of reading impaired at 3rd grade and the assessment of training effects. 149 p. Summary 3 p. 2012.
- 444 KOTONEN, TOMMI, To write a republic - American political poetry from Whitman to 9/11. 121 p. Tiivistelmä 2 p. 2012.
- 445 VASALAMPI, KATI, Appraisals of education-related goals during educational transitions in late adolescence: Consequences for academic well-being and goal attainment. 65 p. (99 p.) Yhteenveto 3 p. 2012.
- 446 HANNULA, MARJA, Dialogia etsimässä - pienryhmäkeskusteluja luokassa. - Seeking dialogue classroom discussions in small groups. 121 p. Summary 3 p. 2012.
- 447 PAKARINEN, EIJA, Relations between teacher-child interactions and children's learning and motivation in Finnish kindergartens. 96 p. (175 p.) Yhteenveto 3 p. 2012.
- 448 RÄIKKÖNEN, EIJA, Is timing everything? A longitudinal perspective on adult transitions, their antecedents, and psychological implications. - Riippuuko kaikki ajoittumisesta? Aikuisuuden siirtymät, niiden ennakoijat ja yhteys psykologiseen toimintakykyyn pitkätaistutkimuksen näkökulmasta. 88 p. (159 p.) Tiivistelmä 5 p. 2012.
- 449 TILLI, JOUNI, The Continuation War as a meta-noic moment. A Burkean reading of lutheran hierocratic rhetoric. 200 p. Tiivistelmä 3 p. 2012.
- 450 LAUTAMO, TIINA, Play Assessment for Group Settings (PAGS): Validating a measurement tool for assessment of children's play performance in the day-care context. 69 p. (114 p.) Yhteenveto 2 p. 2012.