DISSERTATIONES PEDAGOGICAE UNIVERSITATIS TARTUENSIS
11

DISSERTATIONES PEDAGOGICAE UNIVERSITATIS TARTUENSIS
11

MAIRI MÄNNAMAA

Word Guessing Test as a measure of verbal ability. Use of the test in different contexts and groups

TARTU UNIVERSITY
PRESS

Institute of Pedagogy, Faculty of Social Sciences and Education, University of Tartu, Estonia

Dissertation is accepted for the commencement of the Degree of Doctor of Philosophy (in Pedagogy) on August 10, 2010 by the Doctoral Committee of the Faculty of Social Sciences and Education, University of Tartu, Estonia

Supervisor: Eve Kikas, PhD, Professor
University of Tartu
Estonia

Oponents: Pekka Niemi, PhD, Professor
University of Turu
Finland

Marika Veisson, PhD, Professor
University of Tallinn
Estonia

Commencement: October 15, 2010

ISSN 1406–1317
ISBN 978–9949–19–462–9 (trükis)
ISBN 978–9949–19–463–6 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 454

CONTENTS

LIST OF THE ORIGINAL PUBLICATIONS

This dissertation is based on the following original publications which will be referred to in the text by their respective Roman numerals:

I Männamaa, M., Kikas, E., & Raidvee, A. (2008). The effect of testing condition on word guessing in elementary school children. *Journal of Psychoeducational Assessment, 26,* 16–26.

II Kikas, E., Männamaa, M., Kumari, V., & Ulst, T. (2008). The relationships among verbal skills of primary school students with specific learning disabilities and a typically developing comparison group. *International Journal of Disability, Development and Education, 55,* 315–329.

III Männamaa, M., & Kikas, E. (2010). Cognitive profiles and their stability in different academic performance groups of math and language. In A. Toomela (Ed.), *Systemic person oriented study of child development in early primary school* (pp. 95–131). Frankfurt: Peter Lang.

The author of the dissertation contributed to these publications as follows:

- In Study I: working out the instrument for the study, formulating the research questions, carrying out the data collection (individual testing) and parts of the data analysis as well as writing the manuscript;
- In Study II: participating in working out some of the measures used in the study, carrying out statistical analyses and writing the manuscript;
- In Study III: formulating the research questions, participating in data collection, carrying out data analysis and writing the manuscript.

PREFACE

According to the data of the Ministry of Education and Research, Statistics Estonia, Estonian Information System for Education (EHIS) the number of students who leave school before acquiring basic education has shown a raising tendency during the last years (Statistikaamet, 2008). The reasons for dropping out of school are different, including poor academic achievement or misbehaviour. Currently the number of students, who are subject to support systems (e.g., long day groups, speech therapy, remedial groups, individual curriculum) in regular schools because of their special educational needs, has also increased. In case of serious academic underachievement in three or more subjects and if the support systems have not been effective, repeating a grade is designated (see Põhikooli ja gümnaasiumiseadus, 2010). In addition, as the discussions in the media have highlighted – the fulfilment of the requirements of curriculum standards is a big challenge for both teachers and children. The unbalanced proportions of the taught material and the number of lessons have considerable influence on learning outcomes, especially among students with low achievement levels. Among elementary school children, spelling, reading and math difficulties are the most common (about 42%), followed by learning problems in general (about 26%; see Kanep, 2008).

Although there are some positive examples from history and recent years of adapting well-established intelligence tests to the Estonian culture (e.g., Tork adapted a test battery of intelligence and measured the abilities of more than 8000 Estonian schoolchildren and Pullmann standardized the Raven Progressive Matrices; see Pullmann, 2005; Tork, 1940), the main problem in Estonia is the availability of only a few original or adapted measurement tools, which could be used by specialists in educational or medical systems for assessing a child's abilities. Also, there is no unified system or organization for adapting and standardizing well-known tests from other countries. The large amount of work required to adapt and standardize tests in comparison to the small number of potential users of these tests constitutes the main obstacle for creating a unified system (Kikas, 2006; Männamaa & Kikas, 2009). The tests and norms need constant revision and refreshment in order to give an adequate basis for interpreting test scores (AERA, APA, & NCME, 2004; Kanaya, Scullin, & Ceci, 2003). Also, the scores of general ability are not as informative as scores of specific abilities in the practical preventive and intervention work with children who need additional support.

Thus, the combination of increased number of children with special educational needs and limited resources to identify these children as early as possible, has referred to the need to develop a tool which could be used both by specialists in the educational system and in clinics. It is more efficient to deal with prevention of academic failure than working with the consequences.

The aim of the doctoral dissertation was to develop a test of verbal ability – Word Guessing Test (WGT) – which might be used by the educational

professionals, e.g., school psychologists, special teachers, speech therapists and also by clinical psychologists both for screening and intervention purposes for identifying problematic and academically at-risk children. The doctoral dissertation is based on three original publications and an introduction, which includes the theoretical basis of the dissertation (outlining the relations between verbal abilities and academic success and the assessment of children's abilities), overview of the process of the development of the original test, description of the methods used in the empirical studies, and analyses of the findings (psychometric properties, the effect of the testing context and time on WGT performance) based on three empirical studies.

1. THEORETICAL BASIS

1.1. Verbal abilities and academic success

Hundreds of studies have examined the relationships between various cognitive abilities and school success. Although non-verbal abilities will always remain important in the learning process, the most powerful influence on learning outcomes and academic achievement relies on verbal abilities. Depending on the level of analysis, verbal abilities consist of different components, e.g., vocabulary, comprehension, reading, writing and listening skills, knowledge and verbal reasoning. The National Curriculum for Basic Schools and Upper Secondary Schools (Põhikooli ja gümnaasiumi, 2007) states the learning outcomes of different subjects during the first school level and stresses the importance of separate and integrated verbal abilities in achieving these goals, e.g., reading and writing skills, comprehension of oral and written language, listening skills, the knowledge of concepts.

In school, learning is based mostly on verbal explanations and children learn mainly through language, where words are not always directly related to concrete objects or personal experiences (Toomela, 2003; Vygotsky, 1934/1977). The decontextualised use of language is encouraged and stressed in school (Carlisle, Fleming, & Gudbrandsen, 2000; Fukkink, 2005; Kikas, 2003; Luria, 1976, 1979; Toomela, 2003; Vygotsky, 1934/1977). Familiarity with a hierarchical word system and the use of verbal concepts reflects the level and development of language and thinking. Use of definitions presumes both linguistic and metalinguistic knowledge (see Toomela, 2003) and provides indirect evidence of the child's vocabulary and reading skills. Conceptual development is in turn related with cognitive development and general ability.

Studies have shown that verbal ability is a good predictor of academic performance (Durand, Hulme, Larkin, & Snowling, 2005; Lepola, Niemi, Kuikka, & Hannula, 2005; Watson et al., 2003). Conceptual skills have been proven important in solving math word problems (Fuchs et al., 2005, 2006; Passolunghi & Siegel, 2001, 2004; Swanson & Sachse-Lee, 2001) and are related to performance in reading and language tasks (Snow, Cancini, Gonzalez, & Shriberg, 1989). Deficits in these skills and their relations to achievement are well documented in several studies. For example, children with learning difficulties obtain worse results in decontextualized tasks than children without learning problems (Gutiérrez-Clellen & DeCurtis, 1999; Marinellie, & Johnson, 2002). Children with language impairments have been found to perform significantly weaker solving math word problems than their non-impaired peers (Jordan, Levine, & Huttenlocher, 1995).

1.2. Assessment of children's abilities

For assessing abilities (including verbal abilities), different tests and test batteries have been developed. All the well-established and widely used tests in the world are available for educational professionals who have approved their qualification to test selling centres and have enough resources to buy the tests. As mentioned before, Estonia does not have a unified system or organization for adapting and standardizing tests from other languages, taking responsibility for refreshment of norms and assuring that the adaptation process follows high, theory-based scientific standards (e.g., Psykologien Kustannus OY in Finland). For assessing children's abilities the complex and time-consuming test batteries, e.g., Wechsler Intelligence Scale for Children (WISC, Wechsler, 1991, 2005), Kaufman Assessment Battery for Children (K-ABC, Kaufman & Kaufman, 1983, 2004) are used by a limited group of psychologists who mainly work at clinics. Clinics are usually large institutions which provide resources to purchase the well-known tests from other countries and stand for applying these assessment tools in psychologists' everyday work. In school the resources are much more limited. However, the assessment of children's abilities and need for proper tools is necessary and crucial also in school context.

Assessment in school context and also in clinics is mainly targeted towards identifying at-risk children. The interpretation of test results and validity of conclusions based on test scores are directly dependent on the psychometric properties and the correct and updated norms of the test. Interpreting the results in context and the use of correct norms is crucial for assessing children (AERA, 2004; Kline, 1994). Thus, there is detectable need for screening tests that could be used not only in clinical settings, but also in school, tests that are available both to psychologists and educational professionals. Furthermore, the psychometric properties of the tests should be acceptable and meet the standards for educational and psychological testing (AERA, 2004).

Similarly to learning and development (Valsiner, 2000; Vygotsky, 1934/1977), the role of social context should be stressed in assessment. When developing and using the tests for assessing a child's abilities, it is important to keep in mind that the child's performance might be influenced by several factors such as the cultural background, testing context, primary language, child's motivation and emotional state (see AERA, 2004; Kim, Baydar, & Greek, 2003; Oakland, 2009; Sattler, 2001). As demonstrated in earlier studies, retrieval from memory is much more effective when the conditions are similar to those of the learning situation (Chu, Handley, & Cooper, 2003; Smith, 1994; Tulving, 1985). Also, sometimes children may fail to generalize their knowledge (see Boschowitsch, 1968; Watson, 1985) and might give different answers in different contexts (Bjorklund & Rosenblum, 2003). Based on these findings it may be expected that children use their skills and perform better in tests related to academic knowledge in school context. Still, there is some controversy regarding optimal testing conditions. Some studies have revealed a

positive effect of school context and the group-administered testing conditions on children's performance (Barth, Dunlap, Dane, Lochman, & Wells, 2004; Crozier & Hostettler, 2003), whereas others report better test results in individually administered conditions (Hundeide, 1985; Milgram & Milgram, 1976). This also refers to the need to carry out children's assessment in different contexts – at clinics and in school.

2. THE AIMS AND RESEARCH QUESTIONS

The aim of the doctoral dissertation was to develop two versions of the WGT, to examine their psychometric properties and analyse the results of the WGT in different contexts and time points. The final adaptation of the long version consists of 60 items and the short version of 8 items. The long version of the WGT was tested in two testing conditions – individual and collective. From a variety of psychometric properties the following were examined: reliability (internal consistency and test-retest) and three types of validity (construct, concurrent and predictive). The individual differences and stability of the WGT performance over time were assessed using group and person-oriented data analysis methods.

The specific research questions posed and the studies where the answers to the questions were searched for are the following:

1. Examining the reliability of the long (Studies I and II) and short (Study III) versions of the test.
2. Examining the validity of the long (Study I and II) and short (Study III) versions of the test, specifically construct (Studies I–III), predictive and concurrent validity (Studies II and III).
3. Examining the suitability of using the test in different conditions – individual and collective testing (Study I).
4. Comparing the results of WGT in different contexts (Study I).
5. Examining the individual differences and the stability of the results of the WGT inside different level academic performance groups (Study III).

Additionally, other studies related to the aims of the dissertation, in which WGT has been used, are referred to (Kikas, Peets, Palu, & Afanasjev, 2009; Mägi, Häidkind, & Kikas, 2009; Soodla & Kikas, 2010).

3. METHOD

3.1. Background and the development of the WGT

As shown in previous studies, different subskills of verbal abilities are associated with achievement and school success (Durand et al., 2005; Fuchs et al., 2005; 2006; Lepola et al., 2005; Passolunghi & Siegel, 2001, 2004; Snow et al., 1989; Swanson & Sachse-Lee, 2001). Assessing verbal abilities is an ambitious and complicated task. Therefore we delimited the task to one specific sub-skill of verbal abilities – word guessing, and concentrated on the development of the test assessing this particular skill. Word guessing as a construct includes diverse information about different verbal subskills, including vocabulary, comprehension, conceptual skills and verbal reasoning. These subskills, taken separately, are all essential skills necessary to achieve good learning outcomes in academic context.

The main idea of the test is similar to other subtests in intelligence and neuropsychological test batteries, e.g., the Riddles subtest of the Kaufman Assessment Battery for Children (K-ABC, Kaufman & Kaufman, 1983, 2004), Word Reasoning subtest from the new version of the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV, Wechsler, 2005) and The Word Context subtest from Delis-Kaplan Executive Function System (Delis, Kaplan, & Kramer, 2002). These tasks have been shown to assess verbal comprehension, analogical and general reasoning, ability to integrate and synthesize different types of information, verbal abstraction, domain knowledge and the ability to generate alternative concepts.

The Riddles subtest of the K-ABC (see Kaufman & Kaufman, 1983) was taken as a starting point and exact translations of the 32 items of the subtest items from English to Estonian were used. The first findings (571 children, age 2.5–12.5) with the translated items demonstrated that the Riddles subtest of the K-ABC (Kaufman & Kaufman, 1983) did not function satisfactorily for all age groups and items (see Männamaa, 2000). Thus, further modifications were needed to bridge the gap between different languages and cultural environments. Development of an original test assessing word guessing ability in Estonian cultural-language background was of cardinal importance.

3.2. Description of the WGT

In the WGT the child's task is to infer the name of a concrete or abstract concept by three of its given characteristics. The test items are based on the Estonian National Curriculum for different subject areas (see Põhikooli ja gümnaasiumi, 2007) and the majority of the words for identification are provided in elementary school textbooks of different subjects (e.g., math,

language, science). All descriptions include three characteristics of an object, which may be a general category (e.g., "a very old game" for "chess"), a description (e.g., "has jagged edges" for "stamp") or a function (e.g., "helps find direction" for "compass"). As an example, the description of "scissors" is the following: "What we use for cutting, it has two blades, which are sharp." The answer is considered correct only if the child responds with the exact word. If s/he describes the object (e.g., "you use it to buy things" for "money") or gives a wrong word (e.g., "knife" instead of "scissors"), the answer is coded as incorrect. In written form spelling or grammatical errors are not taken into account. Although we have used only correct answers in all the studies of this dissertation, analysis of incorrect answers would be a source of valuable information on eventual cognitive difficulties. If the child responds with the word "mouth" instead of "language" or "hangs on the wall" instead of "calendar", it could have several explanations, e.g., with reference to insufficient attention, deficits in word reasoning skills, poor vocabulary, poor ability to integrate information or absence of the target word from the child's vocabulary. Interpretation of incorrect answers is especially important in practical work with underachievers or in clinical groups.

The first version of WGT included 80 items and was meant to be used in the age range of 3–12 years. In the pilot study with the first version all 32 items from the K-ABC Riddles subtest were used as modified formulations and 48 new items based on different school textbooks were added. Items, that were too difficult or too easy (see Anastasi & Urbina, 1997; Kamphaus & Reynolds, 1987), as well as items that had unequivocal meanings, were excluded. Based on the results of the pilot study, some of the test items (e.g., "island", "friend", "fossil", "siren") were replaced with new ones (e.g., "dictionary", "monastery", "litre"). Separate versions for preschool (ages from 3 to 7) and elementary school children (ages from 7 to 11) were developed, with 40 and 60 items respectively (with 24 items overlapping). In addition, long and short versions of the WGT for elementary school children were developed. The long version of the WGT consists of 60 items, the short version of 8 items. The present doctoral dissertation introduces the results based on ages from 8 to 11.

3.3. Samples, subjects and versions of WGT

The psychometric properties of the WGT were studied in different samples of elementary school children from different districts of Estonia. Home language of all the participants was Estonian. The ratio of girls and boys was nearly balanced at the level of the whole samples (Studies II and III), but varied slightly between grades. In all three studies the analyses were done by grades not by age (see suggestions in AERA, 2004; Alexander & Martin, 2004). Due to the date of birth or age of school entrance (some children start school at the age 8) children's age in different grades overlapped, e.g., both second and third

grade children's age range was between 8 to 10 years (Studies I and II). Two different groups of children were recruited for Studies I and II, respectively: those with and without specific learning disabilities. All of the participants studied according to the National Curriculum for Basic Schools and Upper Secondary Schools (2007). Children without learning disabilities (hereinafter as control) studied in regular classes in state schools and children with specific learning disabilities (hereinafter as LD) either in special schools or special classes for children with learning disabilities. LD had been diagnosed by psychiatrists according to the ICD-10 Classification of Mental and Behavioural Disorders (World Health Organization, 1992) and met the criteria for specific learning disability (F81) – specifically reading disorder (F81.0), disorder of written expression (F81.1) or a mixed disorder (F81.3).

In Study III word guessing ability was compared between different level academic performance groups in math and Estonian language. In this study, word guessing was one of the cognitive abilities assessed in groups with different academic performance levels during two consecutive years in Grade 3 and Grade 4. Children were assigned into low, average and high performance groups relying on their test results in math and Estonian language at two time points. Based on changes in students' academic performance in two consecutive years seven groups in math and seven groups in Estonian language performance were differentiated. In three stable groups of math and Estonian language academic performance level (low, average or high) did not change from Grade 3 to Grade 4; in changed academic performance groups the level of academic performance increased (better performance level in Grade 4) or decreased (better performance level in Grade 3). Two increased performance groups for math and two for Estonian language as well as two decreased performance groups for math and two for Estonian language were formed.

The samples in different studies may be described as follows (see also Table 1):

Study I included 764 students from the second (age 8–10), third (age 8–10) and fourth grades (age 10–11). Five hundred forty-one students belonged to the control group and 232 students into the LD group. About half of the students (387) were tested collectively in the classroom and the other half (377) individually in a separate room.

In **Study II,** the participants were 251 students from Grade 2 (age 8–10 years; 31 boys and 34 girls), Grade 3 (age 9–10 years; 44 boys and 27 girls), and Grade 4 (age 10–11 years; 52 boys and 63 girls). One hundred sixty-three children belonged to the control group (45 second, 39 third and 79 fourth graders) and 88 children to the LD group (20 second, 32 third and 36 fourth graders).

Study III was a longitudinal study and here only children who were tested with the WGT at two time points were included. The sample in Study III included 682 third grade children (49.49% boys and 50.51% girls) from 28 Estonian elementary schools and 44 classes. After a one-year period the children were tested again in the fourth grade. The participants' age range was between 8 to 10 years in both grades (M = 9.07, SD = 0.38 and M = 10.00, SD = 0.34, respectively in Grade 3 and Grade 4). Seven academic performance groups in math and seven academic performance groups in Estonian language were differentiated. For math groups in detail: 36, 360 and 54 children belonged to the three stable groups (low, average and high, respectively), 32 and 56 children to the increased (low and average) and 61 and 73 to the decreased (average and high) groups. For Estonian language groups: 40, 347 and 30 children belonged to the three stable groups (low, average and high), 43 and 37 children to the increased (low and average) and 47 and 113 to the decreased (average and high) groups.

Table 1. Overview of the samples, testing conditions and versions of the WGT in different studies

	N	Control	LD	Age range	Test version	Testing condition
Study I	764	541	232	8–11	Long	Collective/individual
Grade 2		170	61	8–10		
Grade 3		190	91	8–10		
Grade 4		181	80	10–11		
Study II	251	163	88	8–11	Long	Collective
Grade 2		45	20	8–10		
Grade 3		39	32	9–10		
Grade 4		79	36	10–11		
Study III	682	682		8–10	Short	Collective
Grade 3		682		8–10		
Grade 4		682		8–10		

Note. Control = children without diagnosed learning disabilities; LD = children with diagnosed learning disabilities; Test version: long = WGT with 60 items, short = WGT with 8 items; Testing condition: collective = WGT administered collectively in the classroom, individual = WGT administered individually in a separate room.

3.4. Procedure

In all studies informed parental consent was obtained for each child. The parents received letters where the aims of the study were described. Children whose parents did not allow their children to participate in the study were excluded.

In Study I half of the children were tested individually and half of the children collectively. In Studies II and III all the children were tested collectively in the classroom (see also Table 1). In individual testing condition all the students were tested in a separate room (in school or at clinics). The experimenter read aloud the test items one at a time, the students answered orally and the experimenter wrote down the answers. In the collective condition students were tested during one lesson with the whole class together. Similarly to the individual testing, the experimenter read out the descriptions of the items to the whole class and children gave the answers on a sheet of paper.

In Study III the short version of WGT was used together with other verbal tasks and all the tasks were completed within the timeframe of one academic lesson.

4. RESULTS AND DISCUSSION

4.1. Psychometric properties of the WGT

For test developers the most fundamental requirement is to assess the psychometric properties of the test and provide evidence that support the interpretation of the test scores by the proposed users (AERA, 2004; Kurpius & Stafford, 2006). The reliability, as well as different types of validities that are needed for meaningful interpretations of test scores are the main necessary properties of a test. Although the choice of different types of reliability and validity is much wider, reliability of the WGT was estimated by internal consistency and test-retest reliability and validity by construct, predictive and concurrent validity.

4.1.1. Reliability of the WGT

Tests exhibit different kinds of reliability, the importance of which vary on how the test is to be used. Reliability may also vary across different populations (see AERA, 2004; Strauss, Scherman, & Spreen, 2006). The indicators of reliability were good for both the long (Studies I and II) and the short version (Study III) of the WGT, in collective and individual contexts (Study I), and for control as well as LD groups (Studies I and II). Internal consistency of the test (Cronbach α) was at least .85 in all samples. More specifically, for the long version, α was .87, .89, .92 for the control group and .90, .88 and .91 for the LD group (Study I individual condition, Study I collective condition, and Study II, respectively). The internal consistency of the short version was also good: .87 and .85 in Grades 3 and 4 respectively (Study III). All the internal reliability estimates of WGT across the different studies and samples were either good or excellent and refer to the consistency of measurement for making decisions based on assessed test scores.

Test-retest reliability of WGT was not as good as the internal reliability. Correlation between test results from one year to the next (short version, Study III) was .65 showing its marginal stability in time. However, in different academic performance groups, the test-retest correlations varied even from .37 to .71, referring to the variation in WGT scores' stability in different groups of academic performance as well as to the heterogeneity within groups. In addition, the reliability estimates may depend on the type of the assessed ability (e.g., stable or variable) or the time between testing (see Anastasi & Urbina, 1997; Strauss et al., 2006). It is also important to stress that the reliability coefficients do not provide information about the reproducibility of individual test scores or whether or not the individuals retain their relative place in the distribution from the first to the second testing (see Bland & Altman, 1986;

Strauss et al., 2006). To assess the stability of individual test scores person-oriented methods of data analysis are more informative.

4.1.2. Construct validity of the WGT

Good construct validity of WGT was supported by the findings from Studies I–III. First, test scores were higher in upper grades for both conditions (individual and collective) and for both groups (control and LD) (Study I). Specifically, for the control group results of the Scheffé test proved that WGT scores in the collective condition were significantly different between all grades and in the individual condition between the second and other grades. For the LD group the WGT differences in the collective condition were statistically different between the second and higher grades and in the individual condition between the fourth and lower grades. The same findings of grade differences between the WGT scores appeared in Study II, where the main effect of grade was observed, $F(2, 245) = 49.31$, $p < .001$, $\eta^2 = .29$. For the control group the differences were significant between all grades ($p < .001$), whereas for children with LD, differences among grade levels were not significant.

Second, children with learning disabilities performed worse than children in the comparison group. In Study II discriminant analysis revealed that WGT was one of the best predictors for classifying the groups with and without LD, especially in Grade 2 – Wilks'$\Lambda = .85$, $p = .001$. The WGT results in LD groups were significantly lower in all three (second, third, fourth) grades (Study II). The same results of differences between control and LD groups of children were supported by Study I where children with LD obtained lower scores than control children and this effect was detectable for both the collective and the individual testing situations. Specifically, significant main effects of group (control vs LD) were observed, $F(2, 356) = 77.126$, $p < .001$ and $F(1, 232) = 57.087$, $p < .001$, respectively. Differences between control and LD children were significant in all elementary school grades. These results support the findings from earlier studies where deficits in verbal abilities, conceptual skills and verbal reasoning skills of LD groups have been demonstrated (Gutiérrez-Clellen & DeCurtis, 1999; Jordan et al., 1995; Marinellie & Johnson, 2002).

Third, word guessing ability was compared in different academic performance groups in math and Estonian language in Study III. In this study word guessing was one of the cognitive abilities assessed in groups with different academic performance levels during two consecutive years in Grade 3 and Grade 4. Word guessing scores differed in academic performance groups, both for math and Estonian language in Grade 3 and also in Grade 4. The main effect of group for math in Grade 3 and Grade 4 was $F(6, 605) = 28.696$, $p < .001$ and $F(6, 626) = 32.782$, $p < .001$, respectively; and for language $F(6, 599) = 34.519$, $p < .001$ and $F(6, 616) = 50.050$, $p < .001$, respectively. Moreover, as shown in

Study III, WGT was one of the best tests alongside others in differentiating academic performance groups in both grades.

In addition to the classical test theory, WGT's construct validity was estimated in the framework of item response theory (IRT) (Study I), which has been increasingly used in psychological and educational research in recent decades (AERA, 2004). A sample-free measurement of item difficulty parameters was conducted. For the 60-item WGT version, item difficulty and dimensionality for individual and collective assessment conditions were determined according to IRT, specifically with Rasch measurement technique (see Andrich, 1988; Embretson & Reise, 2000), which allows for interval scale measurement of ordinal data. The mathematical treatment of IRT is a function that describes, in probabilistic terms, how a person with a higher standing on a trait is likely to provide a response in a different response category compared to a person with a lower standing on the trait. The probability of success depends on the difference between the ability of the person and the difficulty of the item (Bond & Fox, 2007) – the property of the Rasch model that leads to interval scale measurement.

The adequacy of Rasch model applications and the assumptions for the unidimensionality of the scale was estimated. The chi-square, person separation indices and principal component analysis (PCA) of the residuals was assessed. Additionally differential item functioning (DIF) related to testing situation was controlled. The overall power of test of fit to the model was excellent for both testing conditions (individual and collective) with person separation indexes .90 or higher. Similar results of the test of fit were found for the control and LD groups. The PCA of the residuals showed the amount of variance explained by the residuals to be within acceptable limits for both conditions and groups. Specifically the first components of residuals accounted for 5.20% and 3.72% of the overall variance of the residuals in the individual condition for the control and LD group, respectively. The similar indicators for collective situation were 3.91% and 4.33%. Analysis of DIF showed the non-uniform item functioning across both testing conditions and groups. In sum, the conducted Rasch analysis supports the unidimensionality of the scale and non-uniform differential item functioning across both testing conditions.

4.1.3. Predictive validity of the WGT

The short version of the WGT (with 8 items) has been effectively used as a predictor of children's academic achievement in elementary school (in Grades 1–4) in three studies carried out in Estonia (Kikas et al., 2009; Mägi et al., 2009; Soodla & Kikas, 2010). All of these studies have determined the value of the WGT in predicting achievement in math and Estonian language, thus, supporting the predictive validity of the WGT.

Mägi et al. (2009) explored the extent to which children's achievement behaviours and conceptual knowledge, measured by WGT and the Drawing Test at the beginning of the first grade predicted their subsequent math and literacy performance at the end of Grade 1. Path analysis indicated that individually administered WGT tasks were positive predictors of children's achievement.

A three-year longitudinal study by Kikas et al. (2009) examined the role of individual and contextual factors on the development of math skills. Verbal reasoning skills together with pre-math skills in Grade 1 were found to be positively associated with math achievement three years later, in Grade 3.

Soodla and Kikas (2010) in their study with 633 children (ages 8–10) used WGT as a measure of listening comprehension skills together with measures of working memory and phonological/word decoding skills in Grade 3 to predict reading comprehension skills in Grade 4. Regression analysis indicated that all reading-related measures were significant predictors of reading comprehension skills in Grade 4. The WGT score was the best predictor ($\beta = .309$, $p < .01$) even when the autoregressive effect of reading comprehension in Grade 3 was controlled for.

4.1.4. Concurrent validity of the WGT

Concurrent validity of the WGT is supported by the results from Studies II and III. In Study II, the WGT was used together with a memory test and two additional verbal tests – Word Defining and Categorising/Justifying. In the Word Defining tasks children had to define a concrete or an abstract word (e.g., "a chair", "a poem"). In the Categorisation tasks children had to choose two words that "go together" from a word triad and in Justification tasks they had to analyse the associations between three given words (e.g., "drives, strawberry, opens") as well as find the taxonomic category for two of the words (e.g., "verbs" for "driving and opening"). In the Memory Test children had to memorize 21 semantically related words after 3 minutes of presentation (simultaneously presented in random order). All the verbal tests used in Study II require conceptual knowledge, but the assessed abilities of WGT were better comparable to the Justification Test that requires integration of information as well as memorizing the descriptions (cues).

In the typically developing comparison group the WGT had high and moderate positive correlations with the Justification ($r = .62$) and the Memory Test ($r = .48$), whereas in the LD group the associations between various test results were somewhat different. The respective correlation between WGT and Justification was .42. Unexpectedly, no significant associations between Memory Test performance and WGT were found for the LD group.

Although the Justification Test used in Study II requires the ability to integrate information, it seems that a child's responses in WGT are not as

sensitive to the size of his/her vocabulary as in the Justification Test and the child (also the LD child) could derive the correct word or concept without knowing or understanding all the given cues word for word. These results could be explained by the findings from other studies demonstrating difficulties LD children express in word retrieval, analysing the language or integrating information (see Cain, Oakhill, & Lemmon, 2004; Faust, Dimitrovsky, & Shacht, 2003; Korkman & Pesonen, 1994). As in WGT tasks the child had to memorize the given cues, the association between WGT and the Memory Test seems obvious. As the control group outperformed the LD children, spontaneous use of memory strategies, better reading skills or faster information processing might explain the weaker results of LD children in memory tasks (see Semrud-Clikeman, 2005; Vicari, Pasqualetti, Marotta, & Carlesimo, 1999). Moreover, the developmental changes in children with LD are not as evident as in children without diagnosed LD (Culbertson & Edmonds, 1996).

In Study III, word guessing ability had the strongest correlations with two tests of general ability (Raven and Figure Finding). The correlations between WGT and Raven were .47 in Grade 3 and .50 in Grade 4 and the correlations between WGT and Figure Finding were .46 and .43, respectively. As known from previous studies concerning the Figure Finding Test, language mediates the relations between visual stimuli and spatial relationships, specifically in this test, where children had to find objects that were defined through their spatial positions relative to other objects (Toomela, 2002) and successful performance on these tasks required the integration of knowledge of spatial relationships, understanding of verbal information and ultimate integration of these skills. Raven is a well-known culture-free measure of general ability, specifically a measure of non-verbal and fluid abilities (Kline, 1994; Strauss et al., 2006). The relation between Raven and WGT (the latter presumes verbal reasoning) may suggest that a verbal component is important in solving Raven tasks as well.

Substantially lower correlations were found between WGT and other verbal tests – .17 and .20 with Categorization and .20 and .36 with Definitions, in Grade 3 and Grade 4 respectively. It should be mentioned that, although similar, these were not the same tests that were used in Study II (also called Categorization and Definitions). The tests used in Study III contained fewer items (eight for Categorization and three for Definitions) and the internal consistency of Definitions was not good (.58 and .60 in Grade 3 and Grade 4 respectively). The items in Categorization Test were quite similar in their level of difficulty, despite the good internal consistency estimates (.84 and .83 in Grade 3 and Grade 4). However, it also implies that WGT does not measure verbal ability alone, but to some extent also general ability, reasoning and the integration of verbal and nonverbal information (skills also necessary for Raven and Figure Finding tests).

4.2. Testing in different context and time

Besides a test's psychometric properties an array of factors may influence children's testing performance, including these related with a child's personality or the test takers' characteristics, test administration, testing context (see Sattler, 2001). Testing context means mostly the arrangement and physical properties of the room, but also the straight rules for test administrating. The role of different testing situations (individual or collective) is not usually stated. Also, test developers generally overlook the developmental changes in cognitive abilities and their relations in time. As a fact findings based on group-level analysis do not give enough information neither about changes in the variability range inside the group(s), nor the changes in test performance on an individual level.

The role of context and change as well as variability in test results was examined in Study I and Study III.

4.2.1. The effect of testing situation on performance

To assess the impact of the testing situation on WGT performance, we examined and compared the results in two – collectively and individually administered – conditions (Study I). Students generally scored higher in the classroom-administered collective condition than during individual testing. The effect was tested for both the control and the LD group. In the control group the tendency to get statistically higher scores during classroom testing was apparent in the second, third and fourth grades. A significant difference in LD group's word guessing scores in favour of collective assessing was revealed only in the third grade.

Findings on the effect of the testing situation with verbal tasks have been reported earlier by Crozier and Hostettler (2003). Though differences between testing conditions might be explained by children's emotional states (Crozier & Hostettler, 2003) or motivation (Barth et al., 2004) we did not check the relations between affective states or motivation and test scores in different conditions in our study.

Similar results on the effect of the testing situation on WGT performance were found by Tamm (2006). While in Study I different children were tested in each condition, Tamm examined the performance of the same children in two conditions. Also, based on the present Study I, it is not possible to state that the differences between testing situations are independent of children's general ability level. However, as shown in another study with mainstream pupils (Grades 5 to 7), the effect of the testing condition on word guessing was significant even when the level of general ability was controlled for (Käärt, 2009).

Thus, in accordance with the findings with the same children re-tested (see Tamm, 2006) and the controlled general ability level (Käärt, 2009), the difference between WGT results in collectively and individually administered

testing conditions were revealed. These findings cannot be explained by poor properties of the test either in one or the other assessment condition (see construct validity of the WGT) nor by the level of general ability, but could in fact refer to the effect of the testing situation. It means that different norms should be used for interpreting the scores of WGT in individual and collective testing conditions. Furthermore, the scores of WGT for different groups of children should be interpreted with caution. Based on the findings from Study I it may be presumed that conclusions based on test results (norms) obtained from individual testing most likely will underestimate the performance of children without diagnosed LD in the collective condition, i.e. in the classroom (not only in testing related performance). On the other hand, using test norms that rely on results from children without diagnosed developmental problems in the individual testing condition for screening out at-risk children, one could easily overestimate at-risk children's performance in the collective condition.

4.2.2. Word guessing and individual differences over time

Although results from all three original papers referred to in this dissertation (Studies I–III) confirmed the usefulness of the WGT for identifying at-risk children, need for multiple assessments of different abilities in order to attain more reliable interpretations of test scores and conclusions about children's abilities was demonstrated in Study III.

In this study the profiles of cognitive abilities (including WGT and nine other tests) during two consecutive years (Grade 3 and Grade 4) and stability of these profiles over time in different academic performance groups of math and Estonian language was examined (see Samples, subjects and versions of WGT for detailed descriptions of the academic performance groups). Based on changes in academic performance in two consecutive years, seven groups in math and seven groups in Estonian language performance were differentiated; specifically three stable, two increasing and two decreasing performance groups for both subjects. As shown in Study III, the word guessing ability (also other abilities assessed in the cognitive profile of this study) as well as children's academic performance develops and changes in time. Hence the relations with academic outcomes may also change and do so in different ranges. The latter finding was demonstrated by the variation in word guessing scores in groups with different academic achievement levels. Pearson correlations showed significant associations between the WGT scores in Grade 3 and Grade 4 in six math performance groups of seven (correlations ranging from $r = .55$ to $r = .73$), except in the group of stable low math performance ($r = .10$) and in five Estonian language performance groups (correlations ranging from $r = .53$ to $r = .66$), except in the group of stable low language performers ($r = .19$) and average performers who improved their results in time ($r = .22$). Our findings are in accordance with many earlier studies which refer to the variability and

changes in academic (see Belmont & Belmont, 1978; Juel, 1988; Mooij & Driessen, 2008; Phillips, Norris, Osmond, & Maynard, 2002) and cognitive (Ferrer et al., 2007; Primrose, Fuller, & Littledyke, 2000) performance.

Thus, cognitive abilities and academic attainment are not always stable and therefore it is reasonable to avoid conclusions based upon single and static assessments and rather follow the dynamics and direction of changes in cognitive abilities and academic performance groups. Furthermore, in light of the findings, the educators must pay attention not only to low achievers but also to children whose academic outcomes decrease over time and try to find out the reasons for such a decline.

More detailed analysis of the data showed that low scores in cognitive tests (including WGT) even at two time points do not always refer to low academic performance. Even the stable low, average or high scores did not explain academic performance adequately, i.e. some children achieve average academic outcomes despite of low WGT scores and the academic performance of some children may decrease despite of their high WGT scores. As demonstrated in Study III the group-level findings showed the general trends in academic performance groups, but did not give any detailed information about individuals belonging to the groups. Hence these group-level findings cannot be applied to individuals. For this end the person-oriented analysis could be a reasonable methodological choice. In Study III, the data was analysed on the individual level using Configural Frequency Analysis (CFA). CFA is a person-oriented method that allows examining the individuals or homogeneous subgroups of individuals (see von Eye, 1990; von Eye & Bergman, 2003), specifically types (i.e., observed frequency is significantly larger than could be expected by chance) and antitypes (i.e., observed frequency is significantly smaller than expected by chance). The CFA showed the same combination of cognitive abilities (types, antitypes were not revealed in this study) in different academic performance groups and also the different combinations of cognitive abilities within the same academic performance groups.

With the focus only on word guessing, the results demonstrated that despite low scores in word guessing at both times of assessment, there were more children getting average scores in math or language than could be expected by chance (i.e., types). There were also children, who had high scores in word guessing in Grade 3 and Grade 4, but whose performance in math or language was average. Moreover, analysis at the level of each individual child (not as type) indicated that there were also children in different academic performance groups who were good achievers without high, or despite low cognitive abilities. These data demonstrate the variability and changes in cognitive abilities and academic performance on a more detailed level. The differences between the group-level and the person-oriented analysis suggest the importance of using different methods of data analysis for developing interpretations and wider generalizations based on the findings (Bergman, Magnusson, & El-Khouri, 2003; Toomela, 2008).

5. SUMMARY AND CONCLUSIONS

The main results and conclusions of this dissertation are the following:

1. The internal consistency of the WGT was good or excellent both for the long version with 60 items and the short version with 8 items across the different studies. The same results pertain to different groups (with and without diagnosed LD) and contexts. These findings allow presuming the consistency of the measurement tool for making decisions based on its scores.
2. Test-retest reliability of the short version of the WGT as assessed with a one year interval revealed the test's marginal stability in time. However, there was variation in different academic performance groups. These findings could refer to the heterogeneity within the groups and to possible differences in the change patterns in the assessed abilities. For interpreting the WGT scores over time, more detailed intra-group analysis is needed.
3. WGT scores were higher in upper grades for both conditions (individual and collective) and for both groups (with and without LD). Also, children whose performance was academically better received significantly higher scores in the WGT compared with lower achievers. Thus, the WGT is suitable for differentiating children with learning disabilities. Test scores were significantly lower in the group with LD than in the group without LD. These findings support the construct validity of long and short versions of WGT.
4. Construct validity of WGT (the long version) was supported by the uni-dimensionality of the scale and non-uniform differential item functioning across both testing conditions.
5. The WGT score was a good predictor of academic achievement in elementary school. Findings proved the predictive validity of the WGT short version in predicting achievement in math, Estonian language and reading comprehension skills.
6. WGT scores of the long version were positively correlated with other two verbal tests (Categorization and Justification) and the memory test in the control group (children without diagnosed LD). As the Justification test measures similar constructs (verbal knowledge, integration of information, memorizing the words) assessed also in WGT, these findings supported the concurrent validity of WGT. For LD children the relations between the WGT and the aforementioned tests were somewhat different and could be explained by LD children's difficulties in word retrieval, analysing the language or integrating the information.
7. WGT scores of the short version were also positively correlated with scores of general ability (measured by Raven and Figure Finding) which means that WGT does not measure verbal ability and integration of verbal information alone but to some extent also general ability and reasoning as Raven and the Figure Finding Test do. The unexpectedly low correlations

between the WGT scores (short version) and other verbal tests could be explained by differences in measured abilities or psychometric properties of the latter tests. Correlations with other tests demonstrated the concurrent validity of the WGT.

8. WGT performance may be influenced by the testing context. During collective testing in the classroom children with and without diagnosed LD (control vs LD) got different scores than in the case of individual testing. In the control group the tendency to get statistically higher scores in classroom testing was apparent in all of the studied grades. However, a significant difference in LD group's word guessing scores in favour of the collective assessment was revealed only in the third grade.
9. To interpret the WGT results for a classroom or an individual testing situation, different norms should be used, at least for the long version of the WGT. The scores of the WGT for different groups of children should be interpreted with caution. When assessing children without diagnosed LD in an individual condition one could easily underestimate their performance in the collective condition. Also, screening out at-risk children in the individual testing situation may lead to overestimation of their performance in the collective condition.
10. WGT should be used more than once and the results should be compared over time. Children's cognitive abilities and academic achievement, as well as relations between the abilities and academic outcomes are not always stable and may change and develop differently in time. The WGT scores varied in different groups of academic performance and changed in a different range. These findings suggest that one should rather focus on the dynamics of the assessments and must carefully follow not only the low achievers but also the children whose academic outcomes decrease over time.
11. Once and even twice assessed WGT scores (low, average or high level) did not always refer to similar relations with academic performance. Analysis at the individual level showed that children with similar levels of WGT results could belong to different academic performance groups and children with different levels of WGT results could belong to the same group of academic performance. Thus, children may succeed academically despite their low abilities and their academic performance may decline in spite of high cognitive basis. These findings demonstrate that analysis at the individual level provides different results than analysis at the group level and that group level findings cannot be applied to individuals. The use of different methods of data analysis for developing interpretations and wider generalizations can be suggested.

To sum up, the Word Guessing Test can be used by educational professionals, e.g., school psychologists, speech therapists and special teachers as well as clinical psychologists for identifying children who are at risk of academic

failure. The short version can be used as a screening instrument for the educational professionals in school. The long version can be recommended for clinical psychologists to be used in clinics in addition to other tests. Depending on the purpose, both individually and group administered versions of the Word Guessing Test can be recommended, but different norms must be used. As children's abilities are not always stable over time, the assessment of word guessing ability at different time points is recommendable. In each assessment, one should also consider the individual differences in abilities and utilise both group and person-oriented data analysis methods in interpretation of the test results. It means that children should be assessed in different environments and at different time points using different methods of data analysis and that data should be interpreted in a broader context.

5.1. Limitations

There are also some limitations that need to be addressed. First, before claiming that the testing situation has an impact on children's scores in verbal ability tests, particularly on word guessing scores, current results should be verified using larger samples and data from longitudinal studies. It is also important to consider the fact that the possible effect of the testing context on WGT performance was analysed only for the WGT long version with 60 items. The unidimensionality of the scale and non-uniform differential item functioning across both testing conditions cannot be generalized to versions that consist of fewer or more items.

Second, as the different academic performance groups and LD groups were quite heterogeneous, the results should be generalized with caution.

Third, the possible impact of the assessment form (written vs oral), children's motivation, emotional state or associations with other cognitive abilities (e.g., general ability, memory) was not examined in the study.

5.2. Directions for future

First, future research could focus on the qualitative analysis of the incorrect answers of the WGT. The interpretation of errors may give useful information about the cognitive deficits characteristic to different psychiatric and neurological disorders (e.g., learning and developmental disorders, mental retardation, speech and language disorders, ADHD, epilepsy) and thus prove useful for planning intervention. Perseverations and insufficient integration of information, use of concrete concepts instead of higher order taxonomic categories could demonstrate separate or combined deficits in memory, attention, thinking, and insufficient vocabulary or text comprehension skills.

Second, the effect of the testing context on other tests deserves more attention and should be examined. It is not known whether the effect of testing situation is revealed only upon tests of verbal ability or also upon tests of non-verbal abilities. Answers to this question are necessary for planning assessment but also for developing intervention strategies for at-risk children. Traditionally the assessment of problematic and academically at-risk children is conducted individually (frequently outside of school, in the clinics) whereas the intervention plan is developed in the school and applied in the classroom.

Third, the stability of the effect of the testing situation on performance should be examined in longitudinal studies to find out if there are any age groups (grades) where the effect is more apparent compared to others.

Fourth, the relations between word guessing ability and other cognitive abilities should be analysed in more detail to clarify the underlying cognitive processes' and their combinations' associations with the WGT.

The list of possible directions for the future is not exhaustive. The next practical step would be completing the collection of norm data and preparing the manual for test administration and interpretation.

SUMMARY IN ESTONIAN

Mõistete äraarvamise test verbaalsete võimete hindamiseks. Testi kasutamine erinevates kontekstides ja gruppides

Dissertatsioon põhineb kolmel artiklil ja sissejuhataval osal. Töö sissejuhatus sisaldab teoreetilist osa, milles keskendutakse verbaalsete võimete ja akadeemilise edukuse vahelistele seostele ning laste võimete hindamisele. Sissejuhatavas osas antakse ülevaade mõistete äraarvamise testi väljatöötamise etappidest, kirjeldatakse täpsemalt empiirilistes uuringutes kasutatud meetodeid ja analüüsitakse kolme empiirilise uuringu tulemusi.

Dissertatsiooni autori esmaseks eesmärgiks oli välja töötada mõõtevahend – mõistete äraarvamise test algklasside laste verbaalsete võimete hindamiseks. Kuigi antud metoodika ei võimalda hinnata laste verbaalsete võimete kõiki aspekte (nt verbaalseid teadmisi, sõnavara, arusaamist), eeldab mõistete äraarvamise testi ülesannete lahendamine esitatud verbaalse informatsiooni integreerimist, mõistete tundmist, piisavat sõnavara ja teksti mõistmist. Lisaks annab testiülesannete lahendamine kaudset informatsiooni lapse mälust ja tähelepanuprotsessidest. Võimalikest kognitiivsetest probleemidest saab parema ülevaate testiülesannetele antud vastuste sisu analüüsist, mida selle töö raames aga põhjalikult ei käsitleta.

Mõistete äraarvamise testi väljatöötamisel võeti aluseks sarnase põhimõtte järgi koostatud testid ajamahukates intelligentsust mõõtvates testipatareides. Kuna Eestis on laste võimete hindamiseks kasutatavate mõõtevahendite valik väga napp ja tõlgitud metoodikad ei anna sageli oodatud tulemusi, oli Eesti oludesse sobiv originaalmetoodika väljatöötamine põhjendatud ja vajalik.

Mõistete äraarvamise testi ülesannetes tuleb lapsel esitatud kirjelduste alusel saadud informatsioon integreerida ja leida sellele vastav sõna või mõiste. Testiülesannete koostamisel lähtuti õppekavast ja ülesannetes kasutati neid mõisteid, mis esinesid läbivalt erinevate õppeainete õpikutes. Mõistete äraarvamise testi pikemas lõppversioonis (60 ülesannet) kasutatavad mõisted (I ja II uurimus) on valitud varasemalt veelgi mahukamate versioonidega läbiviidud uurimuste tulemuste põhjal. Mõistete äraarvamise testi lühiversiooni (8 ülesannet) (III uurimus) on oma töödes edukalt kasutatud ka teised uurijad nt lugemisoskuse ja kooliedukuse ennustamisel.

Töö teiseks eesmärgiks oli kontrollida testi psühhomeetrilisi omadusi (reliaabluse ja valiidsuse näitajaid) erinevates testisituatsioonides (grupp ja individuaalne) ja gruppides (õpiraskuste diagnoosiga lapsed ja diagnoosita lapsed). Kolmandaks eesmärgiks oli hinnata testitulemuste stabiilsust ja individuaalseid erinevusi, kasutades nii grupi kui indiviidi tasemel läbiviidud analüüse.

Dissertatsioonis käsitletud uurimisülesanded:

1. Hinnata mõistete äraarvamise testi pikema (I ja II uurimus) ja lühema versiooni (III uurimus) reliaablust.
2. Hinnata testi pikema ja lühema versiooni konstruktvaliidsust (I–III uurimus), ennustavat ja kaasuvat valiidsust (II ja III uurimus).
3. Hinnata testi sobivust kasutamiseks erinevates testisituatsioonides (grupp ja individuaalne) (I uurimus).
4. Võrrelda testi tulemusi erinevates testisituatsioonides (I uurimus).
5. Hinnata testi stabiilsust ja individuaalseid erinevusi erineva akadeemilise võimekusega gruppides (III uurimus).

I uurimuses käsitleti konteksti ja testitingimuste mõju laste mõistete äraarvamise testi tulemustele. Algklasside lapsed (2–4. klass) said klassiruumis kirjalikult läbiviidud testimisel paremaid tulemusi võrreldes individuaalselt ja suuliselt läbiviidud testimisega. Kontrollgrupi laste tulemused (diagnoosita lapsed) olid kollektiivsel testimisel paremad kõikides klassides, õpiraskustega laste grupis olid klassis läbiviidud testi tulemused oluliselt paremad vaid kolmandas klassis. Antud tulemused viitavad vajadusele kasutada erinevates tingimustes läbiviidud testi tulemuste interpreteerimisel erinevaid norme.

II uurimuses võrreldi verbaalseid võimeid õpiraskuste diagnoosiga ja diagnoosita 2–4. klassi lastel. Leiti, et verbaalsetest võimetest, mida hinnati sõnade defineerimise, mõistete äraarvamise, mõistete kategoriseerimise ja põhjendamise ning meeldejätmise ülesannetega, eristavad mõistete äraarvamise testi tulemused nooremaid lapsi vanematest ning õpiraskustega lapsi diagnoosita lastest. Diagnoosita laste grupis olid mõistete äraarvamise testi tulemused põhjendamise ja mälutestiga tugevamalt seotud kui õpiraskuste diagnoosiga laste grupis. Kuigi kõik testid olid head eristamaks õpiraskusteta lapsi õpiraskustega lastest, olid mõistete äraarvamise testi tulemused 2. klassis parimaks õpiraskuste diagnoosiga ja diagnoosita laste grupi klassifikaatoriks.

III uurimuses hinnati laste kognitiivsete võimete profiile ja nende stabiilsust erinevates matemaatika ja eesti keele gruppides kahel järjestikusel aastal, 3. ja 4. klassis.

Vastavalt kahe aasta testitulemuste muutustele eesti keeles ja matemaatikas eristati seitse matemaatika ja seitse eesti keele tulemuste gruppi. Kolmes stabiilses matemaatika ja eesti keele grupis jäid tulemused mõlemal aastal samale tasemele (nõrk, keskmine või tugev), muutuvates gruppides matemaatika või eesti keele testi tulemused teisel aastal, st 4. klassis kas paranesid või langesid. Grupitasemel tehtud analüüsid näitasid, et koos üldist võimekust mõõtvate testidega oli mõistete äraarvamise test parimaks akadeemilise võimekusega gruppide eristajaks mõlemal aastal. Samas näitas konfiguratsiooniline sagedusanalüüs (CFA), et akadeemiliste tulemuste gruppides esineb mitmeid kognitiivsete võimete kombinatsioone ja sarnaseid kognitiivsete

võimete kombinatsioone võib esineda erinevates akadeemiliste tulemuste gruppides. Seega ei näita halvemad või paremad testitulemused mõistete äraarvamise testis (ja ka üldist võimekust hindavates testides) alati akadeemilist edukust või edutust, kuna on ka lapsi, kes saavad suhteliselt häid akadeemilisi tulemusi vaatamata tugevate võimete puudumisele või olenemata nõrkadest võimetest. Tulemused näitasid, et ka diagnoosita laste seas on kognitiivsed võimed varieeruvad ja muutuvad.

Dissertatsiooni põhilised tulemused ja järeldused:

1. Mõistete äraarvamise testi sisereliaablused olid erinevate uuringute tulemustele toetudes head ja väga head nii 60 ülesandega testi pikema versiooni kui 8 ülesandega testi lühema versiooni puhul. Sama kehtib ka erinevates gruppides (õpiraskuste diagnoosiga lapsed ja diagnoosita lapsed) ja erinevates testisituatsioonides (grupp ja individuaalne). Antud tulemused lubavad oletada, et test on kooskõlaline ja testiskooride põhjal võib teha järeldusi.
2. Mõistete äraarvamise testi lühiversiooni aastase intervalliga tehtud kordustest (test-retest) näitas piiripealset stabiilsust, varieerudes erinevates akadeemilise edukusega gruppides. Varieeruvus viitab grupisisesele heterogeensusele või siis hinnatava võime võimalikule muutumisele. Erinevatel ajahetkedel saadud testi tulemuste interpreteerimisel on soovitav keskenduda tulemuste grupisisesele analüüsile.
3. Mõistete äraarvamise testi tulemused olid paremad vanemates klassides erinevates testisituatsioonides (grupp ja individuaalne) ja gruppides (õpiraskuse diagnoosiga ja diagnoosita). Parema akadeemilise võimekusega õpilased said testis oluliselt paremaid tulemusi võrreldes akadeemiliselt nõrgemate õpilastega. Seega on mõistete äraarvamise test sobiv sõelinstrument õpiraskustega laste väljaselgitamisel. Õpiraskustega laste tulemused on kontrollgrupiga (diagnoosita lapsed) võrreldes oluliselt madalamad. Kõik need tulemused toetavad mõistete äraarvamise testi pikema ja lühema versiooni valiidsushinnanguid.
4. Mõistete äraarvamise testi konstruktvaliidsuse näitajaks on ka skaala ühedimensionaalsus ja testiülesannete kallutatuse puudumine ühe või teise testisituatsiooni kasuks.
5. Mõistete äraarvamise test on algklassides heaks akadeemilise edukuse ennustajaks. Uuringute tulemused näitasid, et mõistete äraarvamise testi lühiversioon ennustab edukalt hilisemaid tulemusi matemaatikas, emakeeles ja tekstimõistmisel.
6. Mõistete äraarvamise testi pikema versiooni tulemused olid diagnoosita lastel teistest verbaalsetest testidest (defineerimine, kategoriseerimine ja põhjendamine) positiivselt kõige tugevamini seotud põhjendamise ja mälutesti tulemustega. Kuna nii mõistete äraarvamise test kui põhjendamise test mõõdavad verbaalseid teadmisi, informatsiooni integreerimist ja sõnade meeldejätmist, on need seosed mõistete äraarvamise testi kaasuva valiid-

suse näitajateks. Õpiraskustega laste grupis olid seosed põhjendamise ja mälutestiga mõnevõrra teistsugused ja põhjendatavad varasemate uuringute tulemustega, milles on leitud, et õpiraskustega lastel on raskusi sõnade meenutamisel, kõne analüüsil ja informatsiooni integreerimisel.

7. Mõistete äraarvamise testi lühiversioonis on tulemused samuti positiivselt seotud üldist võimekust mõõtvate testide (Raven, kujundi leidmine) tulemustega. Selle põhjal võib oletada, et mõistete äraarvamise test ei mõõda mitte ainult verbaalseid võimeid ja verbaalse informatsiooni integratsiooni, vaid on sarnaselt Ravenile ja kujundi leidmise testile ka üldise võimekuse ja järeldamisvõime näitajaks. Mõistete äraarvamise testi lühiversiooni tulemuste nõrgad seosed teiste verbaalsete testidega on seletatavad erinevustega mõõdetavates võimetes või testide psühhomeetriliste omadustega. Viidatud seoseid nii selles kui punktis 6 võib pidada mõistete äraarvamise testi kaasuva valiidsuse näitajateks.
8. Mõistete äraarvamise testi tulemusi võib mõjutada hindamise kontekst. Klassisituatsioonis said nii õpiraskuse diagnoosiga kui diagnoosita lapsed individuaalselt läbiviidud testimisega võrreldes erinevaid tulemusi. Diagnoosita laste grupis said lapsed klassis läbiviidud testimisel kõikides klassides (teises, kolmandas ja neljandas) statistiliselt paremaid tulemusi. Õpiraskustega laste grupis oli grupi ja individuaalsituatsioonis saadud tulemuste erinevus oluline vaid kolmandas klassis.
9. Erinevates testisituatsioonides tuleb mõistete äraarvamise testi pikema versiooni tulemuste interpreteerimisel kasutada individuaalse ja grupitesti jaoks eraldi norme. Erinevate gruppide tulemuste interpreteerimisel tuleks olla ettevaatlik. Diagnoosita lapsi individuaalsituatsioonis hinnates on oht nende tulemusi grupisituatsioonis alahinnata. Riskilaste väljaselgitamisel individuaalsituatsioonis võib kergesti nende tulemusi klassisituatsiooni jaoks ülehinnata, vähemalt mõistete äraarvamise testi põhjal.
10. Mõistete äraarvamise testi tuleks kasutada mitu korda ja võrrelda erinevatel ajahetkedel saadud tulemusi. Laste võimed, akadeemiline edukus ja ka võimete ja akadeemilise edukuse vahelised seosed võivad ajas muutuda ja areneda. Mõistete äraarvamise testi tulemused on erinevates akadeemilise edukuse gruppides varieeruvad ja muutuvad ajas erinevalt. Selle põhjal võib oletada, et võimeid (kui ka akadeemilist edukust) tuleks hinnata dünaamikas ja akadeemiliselt nõrgemate laste kõrval pöörata rohkem tähelepanu ka nendele lastele, kelle akadeemilised tulemused aja jooksul langevad.
11. Mõistete äraarvamise testi ühekordse, isegi kahekordse hindamise tulemused (madal, keskmine, kõrge) ei näita alati vastavat seotust akadeemilise edukusega. Indiviidi tasemel tehtud analüüsid näitavad, et erinevates akadeemilise edukusega gruppides on lapsi, kes saavad mõistete äraarvamise testis sarnase tasemega tulemusi ning samas akadeemilise edukusega grupis on lapsi, kelle mõistete äraarvamise testi tulemused on tasemelt erinevad. Lapsed võivad olla akadeemiliselt edukad ka siis, kui

nende võimed on nõrgemad ja nende akadeemilised tulemused võivad langeda vaatamata kõrgele võimete tasemele. Selgus, et indiviidi tasemel läbiviidud analüüsi tulemused erinevad grupi tasemel läbiviidud analüüsi tulemustest ja grupi tasemel leitud tulemusi ei saa alati kasutada indiviidi tasandil leitud tulemuste interpreteerimiseks. Seega on soovitav ja põhjendatud tulemuste interpreteerimiseks ja üldistuste tegemiseks kasutada erinevaid andmeanalüüsi meetodeid.

Töö praktiliseks väljundiks on integreeritud võimeid mõõtva mõistete äraarvamise testi väljatöötamine, mida lühiversioonis võivad skriiningvahendina probleemsete laste väljaselgitamiseks kasutada nii koolipsühholoogid, eripedagoogid kui õpetajad. Sõltuvalt hindamise eesmärgist võib mõistete äraarvamise testi läbi viia nii grupi- kui individuaaltestina, kasutades tulemuste interpreteerimiseks erinevaid norme. Testi pikemat versiooni võib soovitada kasutamiseks kliinilises praktikas. Mõistete äraarvamise testi tulemuste tõlgendamisel tuleb arvestada ka võimete individuaalsete erinevustega ja toetuda erinevatele andmeanalüüsi meetoditele. Mõistete äraarvamise test on vaid üheks verbaalseid võimeid hindavaks vahendiks. Laste võimete hindamisel on oluline arvesse võtta ka teisi lapse sooritust mõjutavaid tegureid, tulemuste interpreteerimisel peaks kindlasti lähtuma laiemast kontekstist ja integreeritud informatsioonist.

ACKNOWLEDGEMENTS

I would like to thank all the people who participated in developing the Word Guessing Test, especially speech therapist Inna Marats from Tartu University Children's Hospital. Together with her we composed the items of WGT. The comments and remarks of Eve Kikas and students of clinical and school psychology in early stage of test development were also helpful in formulating and modifying the test items.

This dissertation was supported by the Grants from Estonian Science Foundation (7388 and 5371), Estonian Ministry of Education and Research (SF0180025s08, 3-2/TA5966), the European Social Fund (via Innove, Measure 1.1, project code 1.0101-0267) and the European Social Fund Programme Eduko (via Archimedes Foundation) Grant 30.2-10.2/1247.

I would like to express my deep gratitude to my supervisor Professor Eve Kikas for her support and patience through all the years of our cooperation. I am also very grateful to Professor Aaro Toomela for the opportunity to participate in the work of his team and to enjoy the inspiring meetings and seminars.

I have to thank my good colleague Inna once again for her continuous support, valuable remarks and advice in developing the test. I am grateful to Krista for the opportunity to share with her the joy and worry during the studies. Thanks also to Kati for helping with the language.

I thank all the co-authors of my studies and especially Aire for her help.

I want to thank all the children who participated in the research. Without support and consent from their teachers and parents the work would have been impossible.

And finally, special thanks to my family, to Ivar, Kaur, Tiiu and Laura. Their support and faith in me was encouraging indeed.

REFERENCES

Alexander, J. R. M., & Martin, F. (2004). The end of the reading age: Grade and age effects in early schooling. *Journal of School Psychology, 42*, 403–416.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2004). *Standards for educational and psychological testing*. American Educational Research Association.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage.

Barth, J. M., Dunlap, S. T., Dane, H., Lochman, & Wells, K. C. (2004). Classroom-environment on aggression, peer relations and academic focus. *Journal of School Psychology, 42,* 115–133.

Belmont, I., & Belmont, L. (1978). Stability or change in reading achievement over time: Developments and educational implications. *Journal of Learning Disabilities, 11,* 80–88.

Bergman, L. R., Magnusson, D., & El-Khouri, B. (2003). *Studying individual development in an interindividual context: Person oriented approach.* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Bjorklund, D. F., & Rosenblum, K. E. (2002). Context effects in children's selection and use of simple arithmetic strategies. *Journal of Cognition and Development, 3,* 225–242.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, i,* 307–310.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Boschowitsch, J. L. (1968). *Litchnost' i jego formirovanije v detskom vozraste.* [Personality and its development in childhood]. Moscow: Izdatel'stvo Moskovskogo Universiteta.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge. *Journal of Educational Psychology, 96,* 671–681.

Carlisle, J. F., Fleming, J. E., & Gudbrandsen, B. (2000). Incidental word learning in science classes. *Contemporary Educational Psychology, 25,* 184–211.

Chu, S., Handley, V., & Cooper, S. (2003). Eliminating context-dependent forgetting: changing contexts can be as effective as reinstating them. *The Psychological Record, 53,* 549–559.

Crozier, W. R., & Hostettler, K. (2003). The influence of shyness on children's test performance. *British Journal of Educational Psychology, 73,* 317–328.

Culbertson, J., & Edmonds, J. (1996). Learning disabilities. In R. Adams, O. Culbertson, & S. Nixon (Eds.), *Neuropsychology for clinical practice: Etiology, assessment, and treatment of common neurological disorders* (pp. 331–408). Washington, DC, US: APA.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2002). *Delis-Kaplan executive function system.* San Antonio, TX: The Psychological Corporation.

Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). Cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *Journal of Experimental Child Psychology, 91,* 113–136.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Faust, M., Dimitrovsky, L., & Shacht, T. (2003). Naming difficulties in children with dyslexia: Application of the tip-of-the-tongue paradigm. *Journal* of *Learning Disabilities*, *36*, 203–215.

Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology,* 43, 1460–1473.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97,* 493–513.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C., & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29–43.

Fukkink, R., G. (2005). Deriving word meaning from written context: A process analysis. *Learning and Instruction, 15*, 23–43.

Gutiérrez-Clellen, V., & DeCurtis, L. (1999). Word definition skills in spanish-speaking children with language impairment. *Communication Disorders Quarterly, 21*, 23–31.

Hundeide, K. (1985). The tacit background of children's judgments. In J. Wertsch (Ed.), *Culture, communication and cognition. Vygotskian perspectives* (pp. 306–332). Cambridge, MA et al.: Cambridge University Press.

Jordan, N. C., Levine, S. C., & Huttenlocher, J. (1995). Calculation abilities in young children with different patterns of cognitive functioning. *Journal of Learning Disabilities, 28,* 53–64.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first to fourth grades. *Journal of Educational Psychology, 80,* 437–447.

Kamphaus, R. W., & Reynolds, C. R. (1987). *Clinical and Research Applications of the K-ABC.* American Guidance Service.

Kanaya, T., Scullin, M. H., & Ceci, S. J. (2003). The Flynn effect and U.S. policies: The impact of rising IQ scores on American society via mental retardation diagnoses. *American Psychologist, 58*, 778–790.

Kanep, H. (2008). Analüüs haridusliku erivajadusega õpilastest ja tugiteenuse kättesaadavusest õpilastele. Uuringu *"Haridusliku erivajadusega õpilane koolikaardil, tugiteenuse kättesaadavusest haridusliku erivajadusega õpilastele. II osa"* Teadus- ja haridusministeerium. Retrieved March 06, 2010, from https://www.hm.ee/index.php?048181

Kaufman, A. S., & Kaufman, N. L. (1983). *K-ABC. Kaufman Assessment Battery for Children. Interpretative manual.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *K-ABC-II. Kaufman Assessment Battery for Children. Second edition. Manual.* Circle Pines, MN: American Guidance Service.

Kikas, E. (2003). Constructing knowledge beyond senses: Worlds too big and small to see. In A. Toomela (Ed.), *Cultural guidance in the development of the human mind* (pp. 211–227). Westport, CT: Ablex.

Kikas, E. (2006). Psychology in Estonia. In S. Jimerson, T. Oakland, & P. Farrell (Eds.), *The handbook of international school psychology* (pp. 91–102). Thousand Oaks, USA: SAGE.

Kikas, E., Männamaa, M., Kumari, V., & Ulst, T. (2008). The relationships among verbal skills of primary school students with specific learning disabilities and a typically developing comparison group. *International Journal of Disability, Development and Education, 55,* 315–329.

Kikas, E., Peets, K., Palu, A., & Afanasjev, J. (2009). The role of individual and contextual factors in the development of maths skills. *Educational Psychology, 29,* 541–560.

Kim, H., Baydar, N., & Greek, A. (2003). Testing conditions influence the race gap in cognition and achievement estimated by household survey data. *Applied Developmental Psychology, 23,* 567–582.

Kline. P. (1994). *The handbook of psychological testing.* London and New York: Routledge.

Korkman, M., & Pesonen, A. (1994). A comparison of neuropsychological test profiles of children with Attention Deficit-Hyperactivity Disorder and/or learning disorder. *Journal of Learning Disabilities*, *27*, 383–392.

Kurpius, S. E., & Stafford, M. E. (2006). *Testing and measurement. A user-friendly guide.* Sage Publications, Inc.

Käärt, K. (2009). *Mõistete ja mälu testide sooritus individuaalsel suulisel testimisel ja kirjalikul grupitestimisel.* Magistritöö. Tartu: Tartu Ülikool.

Lepola, J., Niemi, P., Kuikka, M., & Hannula, M. (2005). Cognitive-linguistic skills and motivation as longitudinal predictors of reading and arithmetic achievement: A follow up study from kindergarten to grade 2. *International Journal of Educational Research, 43*, 250–271.

Luria, A. (1976). *Cognitive development: Its cultural and social foundations.* Cambridge, MA: Harvard University Press.

Luria, A. (1979). *Jazyk i soznanije.* [Language and consciousness]. Moscow: Izdatel'stvo Moskovskogo Universiteta.

Marinellie, S., & Johnson, C. (2002). Definitional skill in school-age children with specific language impairment. *Journal of Communication Disorders, 35*, 241–259.

Milgram, R. M., & Milgram, N. A. (1976). Group versus individual administration in the measurement of creative thinking in gifted and nongifted children. *Child Development, 47,* 563–565.

Mooij, T., & Driessen, G. (2008). Differential ability and attainment in language and arithmetic of Dutch primary school pupils. *The British Journal of Psychological Society, 78,* 491–506.

Mägi., K., Häidkind, P., & Kikas, E. (2010). Performance-approach goals, task-avoidant behaviour and conceptual knowledge as predictors of first graders' school performance. *Educational Psychology, 30,* 89–106.

Männamaa, M. (2000). *Kaufmani testipatarei lastele (K-ABC). Ülevaade testist ja pilootuuringu tulemused.* Magistritöö. Tartu: Tartu Ülikool.

Männamaa, M., & Kikas, E. (2009). Test development and practice in Estonia. *International School Psychology Association: Newsletter Word Go Around, 36,* 11–12.

Männamaa, M., & Kikas, E. (2010). Cognitive profiles in different academic performance groups of math and language. In A. Toomela (Ed.), *Systemic person oriented study of child development in early primary school* (pp. 95–131). Frankfurt: Peter Lang.

Männamaa, M., Kikas, E., & Raidvee, A. (2008). The effect of testing condition on word guessing in elementary school children. *Journal of Psychoeducational Assessment, 26,* 16–26.

Oakland, T. (2009). How universal are test development and use? In E. L. Grigorenko (Ed.), *Multicultural Psychoeducational Assessment,* (pp. 1–40). Springer Publishing Company.

Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with specific arithmetic learning disabilities. *Journal of Experimental Child Psychology, 80,* 44–57.

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology, 8*8, 348–367.

Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology, 94,* 3–13.

Primrose, A. F., Fuller, M., & Littledyke, M. (2000). Verbal reasoning test scores and their stability over time. *Educational Research, 42*, 167–174.

Pullmann, H. (2005). *The development of intelligence and personality traits among Estonian schoolchildren.* Doctoral dissertation. Tartu: University of Tartu.

Põhikooli ja gümnaasiumi riiklik õppekava (2007). [National Curriculum for Basic Schools and Upper Secondary Schools]. *Riigi Teataja I 2007, 40,* 294. Retrieved March 06, 2010, from https://www.riigiteataja.ee/ert/act.jsp?id=12888846

Põhikooli ja gümnaasiumi seadus (2010). [Basic Schools and Upper Secondary Schools Act]. Riigi Teataja I 2010, 22, 208. *Retrieved June 06, 2010, from* https://www.riigiteataja.ee/ert/act.jsp?id=13315249

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego: M. Sattler, Publisher, Inc.

Semrud-Clikeman*,* M. (2005). Neuropsychological aspects for evaluating learning disabilities. *Journal of Learning Disabilities, 38,* 563–568.

Smith, S. M. (1994). Theoretical principles of context-dependent memory. In P. Morris and M. Gruneberg (Eds.), *Aspects of memory (second edition): Theoretical aspects* (pp. 168–185). London, England: Routledge Press.

Snow, C., Cancini, H., Gonzalez, P., & Shriberg, E. (1989). Giving formal definitions: An oral language correlate of school literacy. In D. Bloome (Ed.), *Classrooms and literacy* (pp. 233–249). Norwood, NJ: Ablex.

Soodla, P., & Kikas, E. (2010). Teachers' judgments of students' reading difficulties and factors related to its accuracy. In A.Toomela (Ed.), *Systemic person oriented study of child development in early primary school* (pp. 73–94). Frankfurt: Peter Lang.

Statistikaamet (2008). *Lapsed. Õpilaste heaolu ja tervis.* Tallinn: Statistikaamet.

Strauss, E., Scherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests. Administration, norms, and commentary.* New York: Oxford University Press.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79,* 294–321.

Tamm, M. (2006). *Algklasside verbaalsete võimete hindamine. Individuaalse ja grupitestimise võrdlus.* Magistritöö. Tartu: Tartu Ülikool.

Toomela, A. (2002). Drawing as a verbally mediated activity: A study of relationships between verbal, motor, and visuospatial skills and drawings in children. *International Journal of Behavioural Development, 26*, 234–247.

Toomela, A. (2003). Development of symbol meaning and the emergence of the semiotically mediated mind. In A. Toomela (Ed.), *Cultural guidance in the development of the human mind* (pp. 163–210). Westport, Connecticut & London: Ablex.

Toomela, A. (2008). Variables in psychology: A critique of quantitative psychology. *Integrative Psychology and Behavioral Science, 42,* 245–265.

Tork, J. (1940). *Eesti laste intelligents. Peadgoogiline, psühholoogiline ja sotsioloogiline uurimus.* Tartu: Koolivara.

Tulving, E. (1985). *Mälu.* Tartu: TÜ Kirjastus.

Valsiner, J. (2000). *Culture and human development.* London: Sage.

Vicari, S., Pasqualetti, P., Marotta, L., & Carlesimo, G. (1999). Word-list learning in normally developing children: Effects of semantic organization and retention interval. *Italian Journal of Neurological Sciences, 20*, 119–128.

von Eye, A. (1990). *Introduction to Configural Frequency Analysis: The search for types and antitypes in cross-classifications*. Cambridge, UK: Cambridge University Press.

von Eye, A., & Bergman, L. R. (2003). Research strategies in developmental psychopathology: Dimensional identity and the person-oriented approach. *Development and Psychopathology, 15*, 553–580.

Vygotsky, L. S. (1934/1997). *Thought and language*. Cambridge, MA: MIT Press.

Watson, R. (1985). Towards theory of definition. *Child Language, 12,* 189–197.

Watson, C. S., Kidd, G. R., Horner, D. G., Connell, P. J., Lowther, A., Eddins, D. A., Kruger, G., Goss, A. D., Rainey, B. B., Gospel, M. D., & Watson, B. U. (2003). Sensory, cognitive, and linguistic factors in the early academic performance of elementary school children: The Benton-IU project. *Journal of Learning Disabilities, 36,* 165–197.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children – Third Edition (WISC-III)*. New York: The Psychological Corporation.

Wechsler, D. (2005). *Wechsler Intelligence Scale for Children. Fourth edition.* San Antonio TX: The Psychological Corporation.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: Author.

PUBLICATIONS

CURRICULUM VITAE

Mairi Männamaa

Date of birth: February 28, 1963
Citizenship: Estonian
Address: Institute of Education, Faculty of Social Science and Education
University of Tartu Salme 1A
50103 Tartu, Estonia
Telephone: +372 731 9653
E-mail: Mairi.Mannamaa@ut.ee

Education

2005–2010	University of Tartu, Faculty of Social Science and Education, Institute of Education, Doctoral studies
1998–2000	University of Tartu, Faculty of Social Science, Department of Psychology, Master's studies (clinical psychology); Master of Science (MSc) in psychology
1981–1987	Tartu State University, Faculty of History, Department of Psychology; Diploma in psychology

Professional employment

Since 1993	Children's Clinic of Tartu University Clinics, clinical psychologist
2007–2009	University of Tartu, Curriculum Development Centre, extraordinary researcher
1990–1992	University of Tartu, researcher in school psychology,
1987–1989	Tartu State University, researcher in educational sociology

Membership in professional organizations

The Union of Estonian Child Psychologists
The International School Psychology Association (ISPA)
The Association of Estonian Psychologists

Research interests

- assessment of cognitive abilities, developing the assessment tools
- children with special needs

CURRICULUM VITAE

Mairi Männamaa

Sünniaeg:	28. 02.1963
Kodakondsus:	Eesti
Aadress:	Haridusteaduste instituut, sotsiaal- ja haridusteaduskond Tartu Ülikool Salme 1A 50103 Tartu, Eesti
Telefon:	+372 731 9653
E-mail:	Mairi.Mannamaa@ut.ee

Haridus

2005–2010	Tartu Ülikool, sotsiaal- ja haridusteaduskond, haridusteaduste instituut, doktoriõpe
1998–2000	Tartu Ülikool, sotsiaalteaduskond, psühholoogia osakond, magistriõpe, MSc psühholoogias
1981–1987	Tartu Riiklik Ülikool, ajalooteaduskond, psühholoogia osakond, psühholoogi ja psühholoogiaõpetaja diplom

Teenistuskäik

alates 1993	Sihtasutus Tartu Ülikooli Kliinikumi lastekliinik, kliiniline psühholoog
2007–2009	Tartu Ülikool, haridusuuringute ja õppekavaarenduse keskus, erakorraline teadur,
1990–1992	teadur, koolipsühholoogia labor, Tartu Ülikool
1987–1989	teadur, haridussotsioloogia labor, Tartu Riiklik Ülikool

Kuuluvus erialastesse organsatsioonidesse

Eesti Lastepsühholoogide Ühing
Rahvusvaheline Koolipsühholoogide Assotsiatsioon (ISPA)
Eesti Psühholoogide Liit

Teadustegevus

- kognitiivsete võimete hindamine ja hindamisvahendite väljatöötamine
- erivajadustega lapsed

LIST OF PUBLICATIONS

Männamaa, M., & Kikas, E. (2010). Cognitive profiles and their stability in different academic performance groups of math and language. In A. Toomela (Ed.), *Systemic person oriented study of child development in early primary school* (pp. 95–131). Frankfurt: Peter Lang.

Männamaa, M., & Kikas, E. (2009). Test development and practice in Estonia. *International School Psychology Association: Newsletter Word Go Around, 36,* 11–12.

Puusepp, H., Kall, K., Salomons G. S., Talvik, I., **Männamaa, M.**, Rein, R., Jakobs, C., & Õunap, K (published online January, 2009). The screening of SLC6A8 deficiency among Estonian families 5 with X-linked mental retardation. *Journal of Inherited Metabolic Disease.*

Õunap, K., Puusepp, H., Kall, K., **Männamaa, M.,** Jakobs, C., & Salomons, G. (2008). The screening of creatine transporter defect among Estonian families with X-linked mental retardation. *Journal of Inherited Metabolic Disease, 31,* 70.

Kikas, E., **Männamaa, M.**, Kumari, V., & Ulst, T. (2008). The relationships among verbal skills of primary school students with specific learning disabilities and a typically developing comparison group. *International Journal of Disability, Development and Education, 55,* 315–329.

Männamaa, M., Kikas, E., & Raidvee, A. (2008). The effect of testing condition on word guessing in elementary school children. *Journal of Psychoeducational Assessment, 26,* 16–26.

Toome, L., Varendi, H., Ilgina, O., Jaanson, E., Kaasik, B., Kruustük, K., Kolk, A., Lõivukene, R., Maas, H., Meriste, S., Mägi, M.-L., **Männamaa, M.**, Pakosta, T., Salong, K., Sander, V., Stelmach, T., Tänavsuu, T., & Utsal, Ü. (2008). Riski-vastsündinute jälgimise juhend lapse esimesel ja teisel eluaastal. *Eesti Arst,* 5, 389–403.

Talvik, I., **Männamaa, M.**, Jüri, P., Leito, K., Põder, H., Hämarik, M., Kool, P., & Talvik, T. (2007). Outcome of infants with inflicted traumatic brain injury (Shaken baby Syndrome) in Estonia. *Acta Paediatrica, 96,* 1164–1168.

Essenson, E., Beilmann, A., **Männamaa, M.**, Napa, A., Liivamägi, J., & Talvik, T. (2004). Epilepsia. Frontaalepilepsia käitumuslik fenotüüp lastel. *Eesti Arst, 8,* 535–542.

Talvik, I., Vibo, R., Metsvaht, T., **Männamaa, M.**, Jüri, P., Heidmets, L.-T., Hämarik, M., & Talvik, T. (2002). Raputatud lapse sündroom. *Eesti Arst, 82*, 23–29.

DISSERTATIONES PEDAGOGICAE UNIVERSITATIS TARTUENSIS

1. **Карлеп, Карл.** Обоснование содержания и методики обучения родному языку во вспомогательной школе. Tartu, 1993.
2. **Ots, Loone.** Mitmekultuurilise hariduse õppekomplekt eesti kirjanduse näitel. Tartu, 1999.
3. **Hiie Asser.** Varajane osaline ja täielik keeleimmersioon Eesti muukeelse hariduse mudelitena. Tartu, 2003.
4. **Piret Luik.** Õpitarkvara efektiivsed karakteristikud elektrooniliste õpikute ja drillprogrammide korral. Tartu, 2004.
5. **Merike Kull.** Perceived general and mental health, their socio-economic correlates and relationships with physical activity in fertility-aged women in Estonia. Tartu, 2006.
6. **Merle Taimalu.** Children's fears and coping strategies: a comparative perspective. Tartu, 2007.
7. **Anita Kärner.** Supervision and research training within the professional research community: Seeking new challenges of doctoral education in Estonia. Tartu, 2009.
8. **Marika Padrik.** Word-formation skill in Estonian children with specific language impairment. Tartu, 2010.
9. **Krista Uibu.** Teachers' roles, instructional approaches and teaching practices in the social-cultural context. Tartu, 2010.
10. **Anu Palu.** Algklassiõpilaste matemaatikaalased teadmised, nende areng ja sellega seonduvad tegurid. Tartu, 2010.