UNIVERSITY OF SZEGED
FACULTY OF HUMANITIES AND SOCIAL SCIENCES
DOCTORAL SCHOOL OF EDUCATION
TECHNOLOGIES IN EDUCATION

MOJAHED MOUSA

COMPUTER-BASED ASSESSMENT AND ENHANCEMENT OF INDUCTIVE REASONING SKILLS: A CASE STUDY OF THE EDUCATION SYSTEM DEVELOPMENT IN PALESTINE

PHD DISSERTATION

SUPERVISOR:
GYÖNGYVÉR MOLNÁR, DSC
PROFESSOR OF EDUCATION

SZEGED, HUNGARY, 2020

TABLE OF CONTENTS

Dissertation summary

The present research aims to assess and enhance Palestinian primary school pupils' inductive reasoning thinking skills in their early school age by means of technology. We intended to explore the possibility of applying online tests that already have established psychometric characteristics to assess pupils in regular Palestinian educational practice at the early stages of schooling. First, we explored the feasibility and the applicability of computer-based assessment among young pupils by testing their basic mouse skills. Second, we moved forward by the adaptation and piloting of a computer-based inductive reasoning test, developed in Hungary to find out its applicability in Palestinian educational context. Finally, we adapted and developed an online training program for inductive reasoning further based on the Palestinian school curriculum and run the intervention study. The present empirical research connects important developing areas of educational research and places them in the context of the development of the Palestinian education system: (a) improving the quality of thinking skills in the Palestinian educational context especially when it comes to early age school children, (b) giving more attention to educational assessment in research and practice, which can open the doors to evidence-based educational developments, (c) using the advantages of computer-based testing, e.g. reducing the timeframe and costs of assessment. In the main part of this study, we investigated the effectiveness of an online intervention programme on different samples, that is, on different groups of pupils having different levels of inductive reasoning, having different socio-economic factors, and gender. Due to the pioneering nature of the present research study in the Palestinian educational context, at the end of the dissertation we provide a multitude of recommendations and suggestions for further researches.

ABBREVIATIONS

PPT	Paper-Pencil Tests
PPA	Paper-Based Assessment
PISA	The Programme for International Student Assessment
TIMSS	Trends in International Mathematics and Science Study
IEA	International Association for the Evaluation of Educational Achievement
CBA	Computer-Based Assessment
D&D	Drag-and-Drop
PP	Paper-and-Pencil
CB	Computer-Based
FF	Face-to-Face
ICT	Information and Communication Technology
MoEHE	Ministry of Education and Higher Education
UNRWA	The United Nations Relief and Works Agency for Palestinian Refugees
IR	Inductive Reasoning
eDia	An On-line Assessment Platform, Electronic Diagnostic Assessment
SPSS	Statistical Package for the Social Sciences
Prezi	The Online Presentation Software Website
eLea	Online Training Platforms
IRT	Functions of Classical and Item Response Theory
ANOVA	Analysis of Variance
SEM	The Structural Equation Modelling
CFI	Comparative Fit Index
TLI	Tucker–Lewis Index
RMSEA	Root Mean Square Error of Approximation
PMEHE	The Palestinian Ministry of Education and Higher Education
OECD	The Organisation for Economic Co-operation and Development
N	Number
M	Mean
SD	Standard Deviation
Fig.	Figural
RQ	Research Question

BA	Bachelor Degree Certificate
MA	Master's Degree Certificate
WLSMV	Weighted Least Squares Mean and Variance adjusted
PCBS	The Palestinian Central Bureau for Statistics

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 The context of the study

Inductive reasoning is considered as one of the basic thinking processes (Klauer & Phye, 2008) strongly connected to higher-order thinking skills (Söderqvist et al., 2012; Molnár et al., 2013). The increasing need for assessing and developing thinking skills in daily school context is a strong topic among various educational experts (Adey et al., 2007). Enhancing and supporting this development becomes a main goal for education systems (Bottino et al., 2007). For this reason, pupils need to develop thinking skills in order to effectively deal with the 21st century challenges of modern society using new technologies.

Palestine is developing its education system partly by embedding new modern educational devices to its everyday school activities (see Shihab, 2014; Alhadath, 2016; Shraim, 2018). The Ministry of Education launched several reforms in various educational areas covering teachers, infrastructure, and the curriculum (see Alhadath, 2016). In the Palestinian curriculum higher order thinking skills are visible to an average degree (AbdulKader, 2014), however, its enhancement should be a stronger part of the curriculum (Barbak, 2012). Researchers confirm the importance of thinking skills in general (Yang & Chang, 2013; Yuda, 2011) and inductive reasoning in specific (see Csapó, 1997; Molnár et al., 2013; Söderqvist et al., 2012; Klauer & Phye, 2008; Hamers, De Koning & Sijtsma, 2000; Klauer, 1996; Tomic, 1995; Klauer & Phye, 2008; Molnár et al., 2013; Pellegrino & Glaser, 1982; Ropo, 1987; Molnár et al., 2013). Therefore, it is recommended to embed thinking skills activities in the school curriculum (Molnár, 2011; de Konig, 2000; Resnick, 1987).

1.2 Problem statement

Many studies have highlighted the importance, the influential effect of thinking skills on pupils’ performance (Bisanz, Bisanz, & Korpan, 1994; Csapó, 1997; Csapó, Molnár, & Tóth, 2009; Klauer & Phye, 2008; Molnár, 2011; 2006; Molnár, Greiff, & Csapó, 2013), while others have focused on the computer-based assessment and enhancement of different knowledge domains (Barnes, 2010; Carson, Gillon, & Boustead, 2011; Csapó, Lőrincz, & Molnár, 2012; Csapó, Molnár, & Nagy, 2014;

Molnár, 2011). Results from these studies investigated that the assessment and enhancement of these skills can improve the teaching and learning processes, develop pupils' performance, and the education system in general.

The integration of these two issues, namely the assessing and enhancing of thinking skills using modern educational technologies are in the spotlight of the 21st century. Education systems worldwide are looking forward to use Information and Communication Technologies (ICT) in their educational institutions instead of the traditional assessment tools. In Palestine, there are no studies investigating the level of pupils' inductive reasoning skills. The Palestinian Ministry of Education is working forward to implement new technologies in the assessment process replacing the current paper-based assessment methods. It wants to benefit from the advantages that ICT brings to the teaching and learning processes such as developed feedback mechanisms, increased reliability and validity, and easy to use assessment instruments.

All the theoretical and empirical studies of the present dissertation are focusing on the realisation of these issues in the Palestinian educational context and investigating how assessment and enhancement of inductive reasoning thinking skills can be realised in the Palestinian school system using modern educational technologies. Four studies were carried out to achieve these goals. There were no similar studies done in Palestine previously.

1.3 The structure of the dissertation

The dissertation consists of seven chapters. The first chapter introduces the study in general, the research problem, highlights the main aims, and describes the context of the study. The description in this chapter provides information about the way the dissertation is organised.

The second chapter is a literature review, collecting, organising, analysing, and evaluating the relevant publications regarding the realisation and advantages of technology-based assessment and training, and studies about inductive reasoning, especially assessing and developing pupils' inductive reasoning skills using modern educational technologies. The following topics have been covered: transition from paper-based to computer-based assessment, highlighting its main features, advantages and disadvantages; definition, models, assessment and enhancement possibilities of inductive reasoning and its importance; game-based enhancement of thinking skills; skills which are inevitable to apply computer-based assessment and enhancement in the classroom, e.g. mouse and keyboarding skills. It also includes data collected from several research studies regarding the factors that might influence the results i.e. socio-economic backgrounds, and it also introduces the cross-cultural validation of the instruments in general.

The third chapter highlights the structure of the Palestinian education system and its development in applying ICT in the school system. The fourth chapter presents the research aims and the structure of the empirical studies. The chapter discusses the research questions and the related hypotheses. The fifth chapter describes the research methodology used in the empirical studies. This includes the research design, sampling, the description of the instruments, the procedures of data collection and the type of analysis used.

The sixth chapter contains the four empirical studies. Each study is discussed in detail, following the order of the research questions presented in chapter three. The first study piloted the possibilities of using computer-based testing in Palestinian schools. It investigated the developmental level of mouse skills among 2nd and 3rd year pupils and analysed the applicability of an online test measuring pupils' inductive reasoning using the ICT facilities of the participating schools. The second study validated the results of the first one and aimed to test the applicability of computer-based testing in Palestine by assessing second, third and fourth graders' (age 7–9) inductive reasoning skills. It also aimed to discover background factors, which can influence the applicability of computer-based assessment (CBA) of Palestinian pupils and tested gender differences regarding inductive reasoning. In the third study, beyond extending the age range of the sample, a revised version of the computer-based inductive reasoning test was applied to fourth and fifth graders. The fourth study focused not only on assessment, but enhancement too, by testing the applicability and the effect size of a computer-based training programme in inductive reasoning through tasks embedded in mathematical content for 9–11-year-old pupils (N=118). The theoretical model of the online training was based on Klauer's "Cognitive training for children" concept and his theory of inductive reasoning (Klauer, 1989).

The seventh chapter consists of the conclusions derived from the discussions of the findings in the four studies. It also includes the recommendations and the suggestions for future research derived from the limitations of the studies.

2 COMPUTER-BASED ASSESSMENT AND ENHANCEMENT OF THINKING SKILLS

Computers can be used in education for many purposes, including assessment, e-learning and edutainment. These ways take the chance to grab the benefits technology offers us beyond the traditional methods which are mainly based on “drill and practise” (Brom et al., 2009) without relating the acquisition of skills to authentic activity (Yelland, 2005). Educators and researchers are concerned about the advantages of digital game-based learning in enhancing pupils’ involvement in the learning process, and about the possibility of developing thinking skills, especially higher order thinking skills (Yang & Chang, 2013), which became more and more important in everyday school context (e.g. Adey et al., 2007) in the 21st century.

Yuda (2011) found major positive influence that game applications, i.e. digital education materials have on elementary pupils’ cognitive development beyond mixing enjoyment with education: fostering pupils’ spatial thinking skill which is a necessary skill to problem solving in several contexts (see also Downs & de Souza, 2005).

There are no more doubts that the implementation of technology in education offers several advantages and possibilities. The use of computer games is considered as a modern and attractive way of teaching since traditional tools do not have the desired motivating power anymore (Bottino et al., 2007).

According to the research results regarding development and enhancing of thinking skills (see. e.g. Molnár, 2011; Molnár, Greiff, & Csapó, 2013), the most sensitive period of development falls on the early years of schooling, that is, at this period of time pupils’ inductive reasoning skills can the most significantly and effectively be developed, more efficiently than later. Additionally, early interventions could have significant impact on later school success in higher schoolyears too (Nagy, 2008). Pásztor (2016) presents some factors to be considered to achieve efficient interventions: the spotlight should be directed toward the familiarity of the structure, nature, and development of different abilities, as well as toward the ability of being able to use the available instruments for everyday application in educational practise to assess and foster thinking skills in an efficient way.

The development of pupils’ thinking skills must be among the most important educational tasks (Resnick, 1987; Molnár et al., 2013), especially in the 21st century. This postulate can be recognized worldwide since the education systems aim to enhance and support the development of pupils' thinking skills to the highest possible level (Bottino et al., 2007).

Klauer & Phye (2008) indicated that about a hundred years ago, empirical research was commenced on inductive reasoning on behalf of intelligence research. According to Spearman's research result, the general intelligence factor, g-factor (see also Csapó, 1997) is basically determined by inductive processes. Inductive processes have been defined and identified as central intellectual factors, as reasoning, or fluid intelligence. Klauer et al., (2002) and Csapó (1997) confirmed the relationship between inductive reasoning and intelligence. Inductive reasoning plays a significant role in the acquisition of new knowledge and skills (Goldman & Pellegrino, 1982).

2.1 Technology-based assessment and enhancement

Computerised testing can be used more efficiently than traditional testing methods to assess pupils' knowledge and skills (Adey & Csapó, 2012). On the one hand, it has several advantages over face-to-face or paper-and-pencil testing. On the other hand, the evolution of technology and the labour market resulted in new expectations, the development of the so called 21st century skills, towards the school systems. These new skills are not equal to the educational standards of the 20th century (Mayrath, Clarke-Midura, Robinson, & Schraw, 2012). Mayrath et al. (2012) emphasised the need for change in educational practises, to move from the industrial revolution to the knowledge revolution, and parallel to these changes the need for new assessments. Mayrath et al. (2012, p. 40) highlighted the necessity for urgent changes in educational practice (including policies, tests, teachers, etc.) as follows:

"When I went to school, I did not receive any training or experiences with 21st century skills. There were no rooms for multiparty games that required timely opportunistic communication and negotiation strategies with invisible players. Collaborative problem solving to achieve group goals was not part of our curriculum. I never learned how to manage limited resources and understand tradeoffs between factors with an interactive simulation. We primarily lived in a world of print in books rather than a rich colorful world of visualizations and multimedia. I never was encouraged or taught how to ask deep questions (why, how, what if, so what) and to explore novel hypotheses because all of our curriculum and subject matters was preplanned by the teacher".

They exposed the gap between pupils' experience in everyday school life and the expectations for using modern technologies in learning and be skilled in the most important 21st century skills. E.g. be a critical thinker, good problem solver, and efficient collaborator, be able to generate new knowledge based on the existing school knowledge. Molnár and Csapó (2019c) highlighted that advanced educational systems are already using technology-based assessment with all its advantages independent of the stake of assessment in the everyday educational context. They supported the notion that computer-based assessment can make every dimensions of learning visible (disciplinary, application, and reasoning). The visibility is provided by the several layers and levels of feedback for the teachers and the pupils both.

In the past two decades the largest educational developments happened in the field of assessment from both theoretical and practical point of view, as a result of the qualitative and quantitative developments in the large-scale international assessments. Significant developments happened in the methods of data analysis and data transfer technology, which have been adapted – taking the local characteristics into consideration – in the national assessment and evaluation systems too (Molnár & Csapó, 2019b), significantly improving their effectiveness (Scheuermann & Pereira, 2008). Different feedback mechanisms have been built at every level of assessment from diagnostic to summative, from low-stakes to high-stakes, from national to international assessments. As a result, technology-based educational assessment can be effectively integrated in the school curriculum, it became a feasible reality today.

There are two major ways to improve assessment using technology: firstly, changing the business of assessment (by incorporating technology into specific assessment processes) which is the core process that defines the enterprise and the core processes can work efficiently with technology (Bennett, 2001; Csapó et al., 2012) as developing tests, generating questions, sharing, reviewing or revising items (see Bejar et al., 2013), and scoring even automatically (Williamson et al., 2006). Secondly, technology can improve assessment by changing the substance of assessment (Bennett, 2001), meaning using something innovative rather than the traditional approaches of assessment, and changing the nature of the constructs intended to be tested. The use of technology in assessment influences other sectors in the education systems like curriculum reform and pedagogical innovation (Csapó et al., 2012).

Technology offers extraordinary opportunities to improve the educational assessment practices (Molnár & Csapó, 2019b). By applying technology, the tasks can be closer to real-life tasks, more innovative, having more dynamism, and measuring more complex skills. More reliable and valid tests can be developed and administered with more realistic, application-oriented and authentic testing environments, which cannot be found in traditional face-to-face or paper-based assessments (Beller, 2013; Bennett, 2002; Breiter, Groß, & Stauke, 2013; Bridgeman, 2010; Christakoudis, Androulakis, & Zagouras, 2011; Csapó et al., 2012; Farcot & Latour, 2009; Kikis, 2010; Martin, 2010; Martin & Binkley, 2009; Moe, 2010; Ripley, 2010; van Lent, 2010). As a result, a significant change in the effectiveness of assessments can be detected (Molnár & Csapó, 2019b). Some of the major opportunities that technology can enforce and improve educational assessment practices are listed below:

- Delivering immediate feedback for teachers and learners both (Becker, 2004; Dikli, 2006; Mitchell, Russel, Broomhead, & Aldridge, 2002; Valenti, Neri, & Cucchiarelli, 2003) using automated scoring which is an important basis of successful learning (Molnár & Csapó, 2019c).
- Enhancing the quality of testing – validity, reliability, and objectivity (Csapó et al., 2014; Jurecka & Hartig, 2007; Ridgway & McCusker, 2003).
- Editing and developing tests in several ways (Csapó et al., 2012) – e.g. fix testing, adaptive testing, using automated or semi-automated generated questions.
- Saving time in test administration and data flow (Csapó, Lőrincz, & Molnár, 2012).
- Cutting and reducing costs of test administration (Bennett, 2003; Choi & Tinkler, 2002; Farcot & Latour, 2008; Peak, 2005; Price et al., 2009).
- Increasing pupils' motivation (Meijer, 2010; Sim & Horton, 2005).
- Developing and presenting interactive and dynamic stimuli, innovative item formats, such as multimedia items containing sounds, animation, video, simulation items, 2nd and 3rd generation tasks and tests (Pachler, Daly, Mor, & Mellar, 2010; Strain-Seymour, Way, & Dolan, 2009; Molnár, Greiff, Wüstenberg, & Fischer, 2017).
- Decreasing measurement error by tailoring tests (like adaptive testing), that is, administering tasks which are suitable for the individual characteristics of the

learners via the availability of adaptive test algorithm (Frey, 2007; Jodoin, Zenisky, & Hambleton, 2006).

- Analysing and logging contextual data (e.g. measuring the time spent on a specific task in a test) (Csapó et al., 2014), that is, the possibility of a full accurate control over the presentation of test stimuli.
- Offering rich and well-structured database to the test result in comparison to paper-based testing that make it possible for researchers, teachers or examiners to analyse pupils' movements and behaviour during testing (Molnár & Lőrincz, 2012).
- Involving learners with learning disabilities (i.e. dyslexia by for example applying audio to the tasks instead of written texts; Csapó, Molnár, & Nagy, 2014).

Based on the several advantages technology-based assessment offers for education, it is expected that the major national and international assessments shift from traditional to technology-based assessment in a short period of time (Molnár, & Csapó, 2019b). All the major projects focusing on the feasibility and comparability issues of computer-based testing (Assessment and Teaching of 21st Century Skills – ATC21S, Class of 2020 Action Plan; Griffin, McGaw, & Care, 2012; SETDA, 2008) concluded that computer-based assessment can be the leading factor in the direction of improvements (Csapó et al., 2012; Pearson, 2012; Scheuermann & Björnsson, 2009).

2.1.1 The transition from paper-and-pencil to computer-based assessment

Educational assessment was mainly based on paper-and-pencil (PP) or face-to-face administration a century ago. There is little doubt today that a sizeable percentage of educational assessment has been shifted its administering mode and now they occur via technology, mostly computer-based (CB). However, when computer-based assessment (CBA) replaces paper-and-pencil or face-to-face (FF) testing, several questions arise (e.g. comparability and equivalence issues, mode-effect, validity, level of technology usage, infrastructure, security).

The courage to replace traditional forms of assessment is increasing in favour of computer-based assessment even with the limitations of technology-based assessment (i.e. high costs of the development of such a system especially at the beginning, media effects and technical conditions in schools), and some educational experts are arguing that the education systems sooner or later will start using the new technological assessment methods over the traditional ones (see Molnár, 2011; Csapó, Molnár, & Nagy, 2014; Kozma, 2009). In everyday educational practice, there are different forces and factors motivating the use of technology-based assessment. It can improve the assessment of already established assessment domains (Csapó et al., 2012) or it makes possible to measure constructs that are fundamental in the 21st century (e.g. problem solving, creativity, critical thinking, ICT literacy) , but would be impossible or difficult to measure (Csapó et al., 2014) with traditional means of assessment (e.g., MicroDYN-based assessment of problem solving; see Greiff, Wüstenberg, & Funke, 2012; collaborative problem solving in technology-reach environment; ICT literacy). In another words, it is difficult to edit tasks measuring 21st century skills without the means of technology (Csapó et al., 2014). Thus, computer-based testing is an "innovative" approach to assessment (Thurlow et al., 2010), however, there are still challenges concerning using it with young learners (Csapó, Molnár, & Nagy, 2014; Carson, Gillon, & Boustead, 2011; Choi & Tinkler, 2002). The limitation of paper-and-pencil assessment (PPA) and the demand for assessing new skills increased the interest in developing technology-based assessment systems. The developments in large-scale assessments, such as Program for International Student Assessment (PISA) influenced these developments significantly and as a result made these systems available for everyday use (Csapó et al., 2012).

Computer-based assessment (CBA) becomes more broad and replaces traditional PP testing around the world. Several studies have been conducted about the shifting from paper-based assessment to computer-based assessment in different knowledge and competence domains to explore the possibilities, advantages, and disadvantages of technology-based assessment (Scheuermann & Guimarães Pereira, 2008), to detect delivery mode-effect on pupils' performance (Clariana & Wallace, 2002; Kingston, 2008), to monitor validity and reliability issues of testing, and to map background factors (e.g. computer familiarity; Csapó et al., 2009; Gallagher et al., 2000) that can have an influence on pupils' achievement in a technology-based environment. With time, the differences between PP and CB test performances are well documented. Most of the referring literature focuses on the comparison of the same construct administered in PP and CB environment and indicated that PP and CB testing are comparable. Comparability problems are not an issue any more by higher grade pupils, as computers became more broadly accessible at schools (Way et al., 2006). In addition, several studies have been conducted to evaluate the comparability of CBA and paper-and-pencil scores (e.g. Csapó, Molnár, & Nagy, 2014; Clariana & Wallace, 2002; Csapó, Molnár, & Tóth, 2009), to measure the effect of administration mode and to discuss the possibilities and challenges CBA offers over traditional assessment methods (see section 2.1). These possibilities and advantages cover two main perspectives: practically by giving immediate feedback, innovative item types and accurate control over the test stimuli, and economically by reducing the costs of test administration (Csapó, Molnár, & Nagy, 2014).

Even large-scale assessments, such as PISA organised by OECD (Organisation for Economic Co-operation and Development), TIMSS (Trends in International Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study) organised by IEA (International Association for the Evaluation of Educational Achievement) started the transition from PP to CBA. PISA have finished this process and from 2015 the default platform is CBA. The IEA is still working on this process with the aim of using computer-based assessment entirely in measuring technological literacy to achieve the goal of supporting and encouraging the development of both international and national assessments which embed the use of ICT in their education system.

PISA belongs to the most prominent international large-scale assessments. Since 2000, in every three years it measures reading literacy, scientific literacy, and mathematical literacy of 15 years old pupils. The main aim of this programme is to (1) provide policy makers with insights on how to help pupils learn better, teachers to teach better, and school systems to operate more effectively; (2) support national policies and evidence-based decision-making; (3) contribute to the education's sustainable development, which emphasises quality and equity of learning outcomes for children, young people and adults.[1]

For the first time ever in 2006, the PISA assessment of science included a computer-based test. Again in 2009 and 2012, PISA offered a computer-based test in a specific field. In 2015, PISA released computer-based items and the major fields (reading, mathematics, and science) were assessed using computerised tools. Over half million pupils from 72 countries attended the internationally agreed two-hours test in 2015. However, this test covers several subjects including science, mathematics, reading, collaborative problem solving, and financial literacy. This programme is one of the

[1] https://www.oecd.org/pisa/aboutpisa/pisa-for-development-background.htm

examples of educational assessment programmes that are shifting gradually from paper-pencil to computer-based assessment.

PIRLS initiated in 2016 a simulated internet environment for the computer-based assessment of online reading (Mullis & Martin, 2019). Mullis and Martin (2019) added that PIRLS is going to give more attention to the digital format in the year 2021 by presenting PIRLS reading passages and items via computer. They said that the main aim of that is to motivate pupils and increase operational efficiency through engaging and visually attractive experience. They are also looking forward to administering PIRLS 2021 on the same digital-based environment as ePIRLS 2021.

IEA's TIMSS used paper-and-pencil-based assessment methods to assess pupils' achievement since 1995 (Martin et al. 2016; Mullis et al. 2016). In 2019, they decided to switch into computer-based assessment aiming to operational efficiencies, enhanced measurement capabilities, and extended coverage of the TIMSS assessment frameworks in mathematics and science (Fishbein, Martin, Mullis & Foy, 2018).

These new possibilities in CBA encouraged several scholars to conduct studies in different knowledge and competence domains and run comparison, with so called media studies. These studies rely on several educational tests which aim to find out the test mode effects on pupils' performance (Clariana & Wallace, 2002; Kingston, 2008). For example, Clariana and Wallace (2002) suggested that instructors and institutions must be aware of the possibility of test mode effects and they also suggested to plan for possible test mode effects with the current increase in computer-based assessment. They found out in their study that computer-based testing has an influence on pupils' performance when the computer-based test group outperformed the paper-based test group. Kingston (2008) synthesized the results of many studies about the comparability of computer-administered and paper-administered tests carried out between 1997 and 2007. He indicated that the results of these studies are not consistent, but also argued for the differences in the measurement and statistical sampling issues of the evaluated studies, which might influence the research results. He argued for the rapid changes over recent years on the quality of computer-based test administration systems as well as the computer-experience of pupils.

The differences between paper-pencil and computer based test performance covered several aspects i.e. advantages and disadvantages, validity and reliability, and the effects of background variables (gender, race/ethnicity, and technology-related factors, like computer familiarity; e.g. Csapó et al., 2009; Gallagher et al., 2000; Csapó, Molnár, & Nagy, 2014). Nowadays, these differences have been extensively studied and well documented (Csapó et al., 2014).

The computer-based administering mode improved the reliability, validity, and objectivity of assessments (Csapó et al., 2014) and is expected to reduce the bias between the two genders in the future applications. The computer-based delivery method provided more valid test results since the answers were automatically coded and scored preventing all the influences/errors that might happen when they are recorded and scored manually by the teachers/human experts. To sum up, according to recent media studies, paper-and-pencil and computer-based test results are comparable, and pupils are in favour of computer-based tests rather than the traditional ones (Csapó, Molnár & Nagy, 2014). Computers are becoming more broadly accessible at schools, which also decreases comparability problems (Way et al., 2006).

As a result of these developments, in the last 20 years technology-based assessment, or to be more precise, computerised testing became the most rapidly developing area of education (Csapó et al., 2012). Technology revolutionized all aspects of assessment in order to facilitate data processing and banking as well as vitalizing the testing situation, increasing motivation, and may improve validity (Csapó et al., 2012).

2.1.2 Software features in computer-based supported pupils' development

Computer games have several features that can be used by designing serious games developed for educational purposes to support pupils' development. Bottino et al. (2007) and Csapó et al. (2012) summarised, evaluated, and highlighted some of these features:

1) There are software products which are able to provide clear advices regarding the way to tackle a specific task (Bottino et al., 2007), for example, a software can present animation steps to reach out a specific aim for the task like solving it.
2) Direct feedback. One of the basic features is giving participants immediate (at least right/wrong) feedback. This supports pupils in error comprehension (Werts et al., 2003). The form of the feedback can have different shapes, e.g. visual, audio.
3) Helping pupils in predicting the next moves (Bottino et al., 2007). It works as a motive to encourage pupils to think of current and future steps as presented in Figure

2.1. In the example provided in Figure 2.1. pupils have to figure out the number that should come in the first yellow square and at the same time they have to figure out also the number that should come in the second empty square.

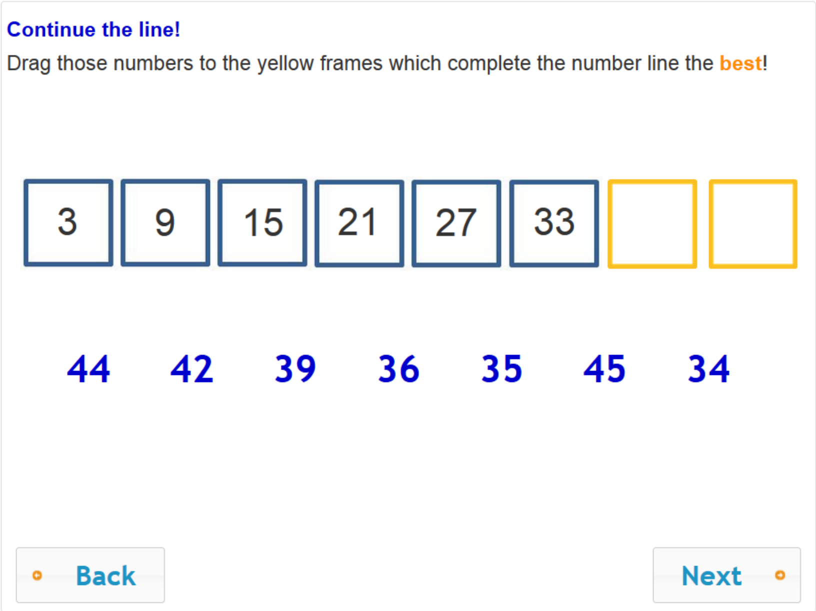


Figure 2.1. Helping pupils in predicting the next moves (task is based on Csapó's (1997) inductive reasoning tasks)

4) Dividing the level of difficulty into degrees, different levels, that is, the level of difficulty could be controlled by the teacher or it can be developed automatically regarding pupils' progress in the tasks (Bottino et al., 2007). It results in a higher motivational level of the pupils as they receive tasks, which difficulty level is close to their ability level and more precise assessment is possible as the level of extracted information about the pupils' ability level is significantly higher than using the same tasks for everybody (see Figure 2.2.; Magyar & Molnár, 2015).
5) Backtracking. Most software programmes give the trainers the possibility (the way differs from one programme to another) to retrace one's step. For instance, some programmes give the user an additional chance to correct, to revise the answer evaluated as wrong. This works also as a feedback to the pupils. This feature can be strongly connected to adaptive testing, where the administration of the tasks, the

difficulty level of the items is strongly depending on the answers given to the previous items (Csapó et al., 2012).

6) It helps in memorising specific actions like review previous moves and visualize some elements related to the upcoming moves (Bottino et al., 2007).
7) Some software products can present specific tips regarding the next move on request by the user (Bottino et al., 2007).

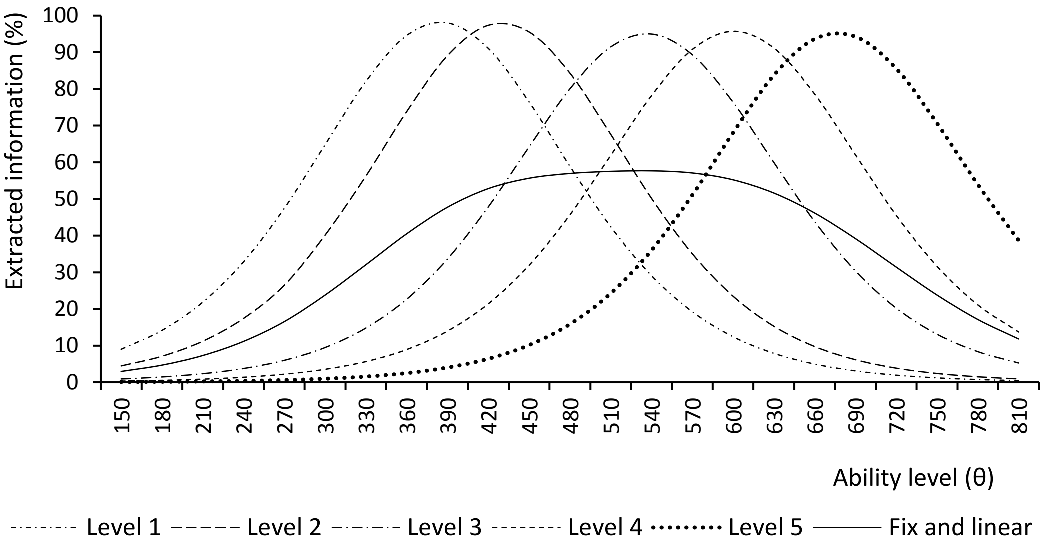


Figure 2.2. Comparing the information curves of linear and multistage adaptive techniques (source: Magyar & Molnár (2015, p. 411))

Most of these characteristics can be found in both technology-based assessments and digital game-based learning such as direct feedback, backtracking, motivational environment, possibilities for interaction within the tasks, importance of optimal challenge and adaptive testing serve the theoretical basis for this endeavour (see Pásztor, 2016). This makes unification possible (Csapó et al., 2012; Pásztor, 2015), that is, using game-based methods embedded in technology-based assessments.

These computer activities do not have negative effect on the pupils' behaviour, but they can work as an incentive for the children/ pupils to cooperate and work together (Brooker & Siraj-Blatchford, 2002). When pupils collaborate, that is, work together on a task, they support one another during the learning process and benefit from the collaboration. Brooker and Siraj-Blatchford argued that "the manipulation of symbols and images on the computer screen represents a new form of symbolic play, in which the children treat the screen images as "concretely" as they do the manipulation of any alternative blocks and small world toys" (Brooker & Siraj-Blatchford, 2002, p. 269). Pupils are engaged when participating in new learning activities reflecting on the relations among symbols, signs, and their relationships. These new technologies drive pupils to use technology in a variety of learning contexts (Yelland, 2005).

2.1.3 An online assessment system for teaching and learning integrating the advantages of CBA: The eDia system

Educational researchers are focusing among others on learning and teaching methods and processes, on the development of skills and competencies using systematic collection and analysis of data collected in educational contexts (see Adey & Shayer, 1994; Shayer, 1999; Adey et al., 2001; Shayer & Adey, 2002; Shayer & Adhami, 2007) to understand the processes of learning, to detect the cognitive and affective development of pupils and provide new methods to make learning and teaching more efficient. Hattie (2009) evaluated the results of more than eight hundred meta-analyses in the field of educational assessment, particularly teaching and learning to find the most influential factors that make teaching and learning successful, to make teaching and learning visible. According to his findings learning "occurs when learning is the explicit and transparent goal, when it is appropriately challenging, and when the teacher and the student both (in their various ways) seek to ascertain whether and to what degree the challenging goal is attained, when there is deliberate practice aimed at attaining mastery of the goal, when there is feedback given and sought, and when there are active, passionate, and engaging people (teachers, students, peers, and so on) participating in the act of learning" (Hattie, 2012, p. 18). According to his research results feedback in all forms, including self-evaluation, self-monitoring, and self-assessment and involvement in the teaching and learning process for both teachers and pupils belong to the most influential factors. He also highlighted that a proper, meaningful, and useful feedback cannot be easily reached. "Learners can be so different, making it difficult for a teacher to achieve such teaching acts: students can be in different learning places at various times, using a multiplicity of unique learning strategies, meeting different and appropriately challenging goals" (Hattie, 2012, p. 18).

Molnár and Csapó (2019c) confirmed that “one of the main challenges of school education stems from the fact that students are different” and they are different in a number of ways, which differences are changing dynamically over time. These differences are not exactly visible, as for example differences referring to height or weight. These cognitive and affective differences are hidden, not visible and are also not easy to make visible, however, Ausubel pointed out already in the 60s, (1968, p. vi) that “the most important single factor influencing learning is what the learner already knows.”

Making thinking visible requires a bunch of teaching and learning practices as Ritchhart et al. (2011) illustrated: “When we demystify the thinking and learning processes, we provide models for pupils of what it means to engage with ideas, to think, and to learn. In doing so, we dispel the myth that learning is just a matter of committing the information in the textbook to one’s memory” (Ritchhart et al., 2011, p. 28). The key problems are memorisation which represents a concept of learning, and information which represents knowledge (Molnár & Csapó, 2019c).

The acronym eDia (stands for electronic Diagnostic assessment) is a system that supports both the learning and teaching processes and makes the three widely accepted dimension of learning (expertise, application, and reasoning; (Csapó & Csépe, 2012; Csapó, Steklács, & Molnár, 2015, Csapó, Csíkos, & Molnár, 2015; Csapó, Korom, & Molnár, 2015; Csapó & Szabó, 2012; Csapó & Szendrei, 2011) visible for the learners and teachers both.

The development of the online assessment system started in April 2007 with the implementation of the TAO open source software (Plichart et al., 2004) at the University of Szeged. The Research Group on Learning and Instruction at the University of Szeged organised several pilot studies with TAO, and collected information, several questions, and problems, which could be raised by developing an online assessment system. At that time, several media effect studies have been also conducted to compare the validity, reliability, and objectivity of paper-and-pencil and online administration (see e.g. Csapó, Molnár, & R. Tóth, 2009). The developmental aim of the TAO platform differed strongly from the main aims of the Research Group, that is, it became obvious that the platform developed at the University of Luxembourg is not suitable and designed for diagnostic assessments. This led to a decision to develop a brand-new platform, which has been optimized for the complex requirements of the diagnostic assessments (Csapó & Molnár, 2019).

eDia is both an online assessment system with almost 20.000 tasks from the domains of mathematics, reading and science and an online assessment platform, suitable for administering first, second and even third generation tasks (Molnár & Csapó, 2019a). The Hungarian technology-based and integrated assessment system uses automatic scoring, that is, it provides immediate feedback for pupils and teachers both (Csapó & Molnár, 2019). The feedback for teachers is elaborated on their pupils' level of knowledge and skills, it is IRT scale-based and also norm referenced, not just a simple achievement-based feedback (Molnár & Csapó, 2019a). Beyond pupil-level feedback it contains a contextualized picture about the class-level, school-level, school-district-level, region-level achievements compared to the country-level mean achievement. Teachers can upload feedback documents for each pupil describing his or her knowledge level in numbers and web figures also, moreover the document contains a detailed text-based description of the pupils' knowledge and skill level too, indicating their weaknesses and strengths (Molnár & Csapó, 2019a).

The system was built to serve the needs of educational practice (Csapó & Molnár, 2019). Csapó and Molnár (2019) explained that the system has other features next to automatic scoring and instant feedback including:

- Item writing which enables editing the items online. Tests can be modified to any language, content and in any format to be suitable for a specific feature of language i.e. direction.
- Test editing which enables constructing adaptive testing techniques that enhance test functionality in minimizing the difference between the pupils' ability level and the difficulty level of the test, based on the given answers.
- Online test delivery. The system can be used anywhere around the world if the person has access link to the system, thus, it is available for everyone everywhere and that makes it a handy tool for researchers, educators to benefit from the characteristics of the system in an easily accessible way.
- Built-in data processing and statistical analyses are important features in the eDia system which makes it possible to make computations like descriptive statistics, classical test theory and IRT modelling for the aim of meeting the assessment needs.

- The ability to make teacher-assembled tests by giving access to the item banks, so teachers can create tests out of the already available tasks in the system, then generate tests to assess pupils individually, or in groups and classes.

Pásztor (2016) used the eDia platform for administering his training program with the aim to enhance pupils' inductive reasoning skills. As a result of his intervention study, he concluded, that teachers have everything they need regarding the results between their hands on one platform, so they do not have to wait for the results as it is the case with the traditional assessment methods. With the eDia system it is possible to carry out large-scale technology-based trainings. He considered the system a knowledge-transformer platform that works as a facilitator among teachers and educational researchers by transferring knowledge among them.

The eDia system can be used for all school stages as well as for university students. It also supports features which makes the testing of early age pupils possible, that is, multimedia elements, like audio or video instructions or game-based, figurative (drawings, figures, pictures) elements can also be used beyond the more traditional letter or number-based characters. The tasks can be administered on desktop computers or mobile devices with touch-screen technologies (e.g. tablets, smart phones) equipped with an internet browser.

2.1.4 The importance of mouse and keyboarding skills in computer-based testing

Administering computer-based tests to young children may raise numerous questions, e.g. regarding pupils' basic computer skills, such as keyboarding and mouse skills, with regard to the feasibility of the assessment and validity of results (Csapó, Molnár, & Nagy, 2014; Barnes, 2010). Despite the widespread and increasing use of computer-based testing even for large-scale assessments, only a few studies have focused on testing very young learners' basic computer skills (keyboarding and mouse skills) in a technology-based environment. Molnár and Pásztor (2015a) explored the potential of using computer-based tests in regular educational practice for the assessment of pupils in the beginning of their schooling, in a study involving almost 5000 first-graders. They distinguished operations based exclusively on mouse clicks, on drag-and-drop and on typing. According to their results operations based exclusively on single mouse clicks proved to be the easiest to perform, it was followed by items consisting typing 1 to 5 numbers or letters only, finally, drag-and-drop (D&D) operations proved to be the hardest, but still executable for most of the pupils. The size and amount of the objects pupils had to click on or drag-and-drop influenced significantly the success and difficulty of the given operation. They concluded that "computer-based assessment and enhancement can be carried out even at the very beginning of schooling without any modern touch screen technology, with normal desktop computers. Most suggested is to use item types requiring mouse clicks and less suggested to use drag-and-drop items." (Molnár & Pásztor, 2015a, p. 118).

2.1.5 Game-based enhancement of thinking skills

The new digital game-based technologies in the field of education add another taste to learning, instruction, and assessment (Kim & Ifenthaler, 2019). The new generation of pupils should be familiar with the twenty-first century demands by teaching them some important things like being creative, innovative, and adaptable, so as to make them prepared to deal with the complex demands of learning in domains (Gee, 2003; Ifenthaler, Eseryel, & Ge, 2012; Prensky, 2001; Shaffer, 2006).

Game-based learning has been identified as a form of pupil-centred learning that places problem solving scenarios within the context of play (Ebner & Holzinger, 2007). Also, it encourages active learning and evokes satisfaction and engagement (Yang & Chang, 2013). Some researchers argue that implementing assessment features into game-based learning environments is still at its beginning times because of many factors like the long-needed process of time to create the design (see Ifenthaler et al., 2012). Prensky (2007) presents several sectors that digital game-based learning interferes in:

- It implements a media-enhanced narrative for creating interest and fosters pupil involvement.
- It is easier and more motivating to complete the tasks with clear instructions provided.
- It integrates interaction and immediate feedback.
- It offers the potential for adaptive learning based on pupils' knowledge and skill level.

- Even the enhancement of the so called 21st century skills like problem solving or creativity is possible.

Based on the primary function of the digital games we can distinguish between games developed for entertainment, and digital games developed for teaching and learning. The latter one is often called serious game. Entertainment, fun, and reaction are the main aims of digital commercial games, while learning and behaviour change are the basic goals of serious games (Connolly et al., 2012). Serious games and game-based learning are being used for the same function and they can be considered as synonyms (Corti, 2006). However, the link between playing games and learning is increasing, and several models which were developed indicated that playing digital games can have a clear learning outcome (Connolly et al., 2012). Garris et al. (2002) divided skills-based learning outcomes (including technical and motor skills), cognitive outcomes (including declarative, procedural, and strategic knowledge), and affective outcomes including beliefs or attitudes which help players to learn by changing their emotions.

Game-based learning has several benefits including stimulating learning motivation and enhancing the interaction between the learning material and pupils (Chen & Huang, 2013). Digital games, for instance, give players the possibility to construct their own understanding naturally (Dormann & Biddle, 2006). It was argued that they can be considered as helpful learning tools (Iacovides et al., 2012). Also, they exercise positive effects on pupils' learning (Pivec, 2007). These positive effects cover two principal areas: learning effectiveness and learning motivation (Chen & Huang, 2013).

Meluso and her colleagues (2012) investigated the effects of game-based learning in the field of science self-efficacy and science content learning. According to her research results the development of pupils' knowledge in science has been noticed, significantly increased (Meluso et al., 2012). Cheng and Su (2012) confirmed this result, that is, the self-efficacy of pupils increased significantly more in a game-based learning environment than in a traditional learning environment. Yang (2012) explored the relationship between pupils' achievements and digital game-based learning. The results were quite positive regarding the enhancement of learning motivation and improvement in promoting pupils' problem-solving skills.

Over the past decades, the development and spread of technological tools offered new training possibilities and methods in game-based environments (Molnár & Lőrincz, 2012). Many scholars supported the idea of computer games playing a significant role in education (Egenfeldt-Nielsen, 2007). Several quantitative studies have been run to explore the educational potential of computer games in education. According to Levin (1981) computer games engage children successfully in teaching math concepts as well as the different ways of approaching math, which is beneficial for individual differences. Klawe (1998) supported the idea that the most effective way to teach pupils math can be attained by computer games. Sedighian and Sedighian (1996) confirmed that computer games are highly effective and influence positively the learning outcomes. According to Betz (1995), Adams (1998), and Rosas et al. (2003) pupils' motivation and learning outcome increased significantly by using computer games as compared to traditional methods. Becker (2001) also supported the stronger correlation between motivation and computer games than between motivation and traditional teaching methods. Pásztor (2014) highlighted that game-based learning environments are handy, where there is no need for the presence of a teacher in a classroom and it can be applied in large groups.

Computer games are promising in many domains: Yelland (2005) stated that they can develop pupils' thinking skills. Molnár (2011) highlighted the power of playful learning environments in developing pupils' reasoning skills. She supported that games provide an effective learning environment for pupils to develop their knowledge and skills. In her experimental study all the basic structures of inductive reasoning: generalization, discrimination, cross-classification, recognizing relations, discriminating between relations, and system formation have been significantly and effectively developed. Pásztor (2014) confirmed this result regarding his online game-based training program developed for pupils to enhance their inductive reasoning skills through tasks embedded in mathematical content.

2.2 Inductive reasoning: definitions, assessment, and enhancement

Inductive reasoning (IR) (Pellegrino & Glaser, 1982; Ropo, 1987; Molnár et al., 2013) belongs to the basic thinking processes (Klauer & Phye, 2008; Molnár et al., 2013). It has connection to almost all higher order thinking skills (Csapó, 1997; Molnár et al., 2013; Söderqvist et al., 2012) like general intelligence (Klauer & Phye, 2008), knowledge acquisition and application (Hamers, De Koning, & Sijtsma, 2000), analogical reasoning (Goswami, 1991), and problem solving (Klauer, 1996; Tomic, 1995). Several studies were established based on these thinking processes. "The inductive method, or teaching by examples, is one of the oldest methods of instruction" (Csapó, 1997, p. 610). Csapó (1997) also added that it is considered as a long-lasting or continuous philosophical problem.

The Multimedia Grolier's Encyclopedia (1994, p. 287) defined inductive reasoning as follows: "induction is a major kind of reasoning process in which a conclusion is drawn from particular cases. It is usually contrasted with deduction, the reasoning process in which the conclusion logically follows from the premises, and in which the conclusion has to be true if the premises are true. In inductive reasoning, on the contrary, there is no logical movement from premises to conclusion. The premises constitute good reasons for accepting the conclusion". The explanation interprets the possibility for the premises to be true and the conclusion to be false since the logical relationship between premises and conclusion are not essential. This gives the importance to induction in the research discovery field since inductive reasoning is used in three main cases: generating hypotheses, formulating theories, and discovering relationships.

There is no universally accepted definition of IR, even though several definitions exist (e.g., Klauer, 1990; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Sloman, 1993; Gick & Holyoak, 1983). Induction is described as "the process whereby one generalizes across a limited number of instances, examples, or observations in order to find a description that applies to them all" (Tomic, 1995, p. 484). Klauer (1989) also defines inductive reasoning as a general principle derived from specific examples.

There are many reasons for emphasizing the importance of induction. Inductive reasoning processes are used in our daily reasoning. We practice this kind of reasoning in our daily life activities in order to reach a specific and certain conclusion; for instance, in expecting the taste of a meal, weather it is going to be tasty or not (Heit, 2001). According to Heit (2001) the second reason for studying induction is that induction is an activity which has many cognitive facets. He provides various kinds of examples regarding this reason; saying that someone could easily give a group of pupils some easy questions and these questions can be delivered to them by using cartoon pictures. In the case of adults, one can provide them with several arguments and their mission is to reach a reason-based judgment. He added that it is not possible to reach a confirmed result since induction itself is not absolute by nature. That has been proven by several studies set by researchers where they are still finding new results. According to Tomic (1995) induction is a helpful procedure to make predictions about new possibilities in order to predict results, as it were. Induction is connected to several types of cognitive activities including decision making, categorization, and probability, as well as similarity judgments (Hayes et al., 2010). Finally, according to Heit (2001, p. 1) “the study of induction has the potential to be theoretically revealing. Because so much of people's reasoning is actually inductive reasoning, and because there is such a rich data set associated with induction, and because induction is related to other central cognitive activities, it is possible to find out a lot about not only reasoning, but also cognition more generally by studying induction”.

According to the classical interpretation, IR covers processes of moving from the specific to the general (direction), that is, generalizing, deriving broad rules based on single experiences and observations (Sandberg & McCullough, 2010). It is a form of reasoning under uncertainty (level of confidence) because it involves forming hypotheses about rules (Perret, 2015). For this reason, it is frequently defined in opposition to deductive reasoning, which involves moving from the general to the particular (direction) and is logical (level of confidence). Other scholars added that inductive reasoning “is described as the generalization of single observations and experiences in order to reach general conclusions or derive broad rules-rule induction” (Molnár, Greiff & Csapó, 2013, p. 36). Induction enables inference with the unobserved, formulates novel conclusions about the unknown, and generates new knowledge (Sloman & Lagnado, 2005). These definitions present the diverse variety among scholars in defining induction.

The most elaborated definition was probably published by Klauer (1993), who interpreted IR as the discovery of regularities through the detection of similarities, dissimilarities, or a combination of both, with respect to attributes or relations to or between objects. This results in a total of six operations of IR: generalisation, discrimination, cross-classification, recognising relations, discriminating between relations, and system formation. It is a helpful procedure to make generalisations about hypotheses, or find out regularities and rules (Klauer, 1993; Klauer et al., 2002).

2.2.1 Directions for teaching and fostering inductive reasoning skills

Teaching thinking skills to children is not an easy mission. There are researchers, who even revise its possibility and argue that it is not possible to make children better thinkers (Jensen, 1973). Nowadays, the importance of developing thinking skills are often discussed in connection with the rapid change of the society and modern technology

“Teaching children to become effective thinkers is increasingly recognized as an immediate goal of education....If students are to function successfully in a highly technical society, then they must be equipped with lifelong learning and thinking skills necessary to acquire and process information in an ever-changing world” (Robinson, 1987, p. 16).

Contrary to early views which rejected the trainability of reasoning skills and emphasised the role of inheritance (Jensen, 1973), empirical studies have indicated that these skills develop over time, mostly in compulsory schooling (Molnár et al., 2013). Despite the relatively slow pace of development, about one-quarter of a standard deviation per year (Molnár et al., 2013), reasoning skills are trainable. Modifiability offers opportunities and new prospects for enhancement through educational interventions (Adey, Csapó, Demetriou, Hautamaki, & Shayer, 2007). Molnár (2011) suggested that thinking skills including inductive reasoning can be significantly and effectively developed by explicit training. Other researchers suggested a different way to achieve the development, like modifying teaching methods (Adey & Shayer, 1994; Shayer & Adey, 2002).

The development of IR appears to emerge at a fairly young age (Perret, 2015), but IR becomes noticeably more efficient with age. According to Csapó (1997), Molnár and Csapó (2011), and Molnár et al. (2013), IR develops over a broad age range, covering the whole period of primary and secondary education. The average pace of development is relatively slow, at about one-quarter of a standard deviation per year (Molnár et al., 2013) resulting from the lack of direct and explicit stimulation of IR in schools. The development occurs spontaneously as a ‘by-product’ of teaching, rather than being guided by explicit instruction (de Konig, 2000; Molnár & Csapó, 2019a).

There are several models of teaching thinking that can be used. Two main approaches are enrichment and infusion. The enrichment approach is mainly based on a specific cognitive theory (McGuinness, 2007). This approach features pre-designed lessons that can be taught next to existing lessons. There are some examples on this approach including the cognitive acceleration programmes (Adey & Shayer, 1994; Shayer & Adey, 2002) and instrumental enrichment (McGuinness, 2007). The cognitive acceleration programmes cover several domains next to the original one which was for science, some other researchers developed this type of programmes for other domains like mathematics (Shayer & Adhami, 2007, 2010), and technology (Backwell & Hamaker, 2004).

The infusion approach on the other hand is a strategy that is based on placing thinking in the context of normal curricular topics aiming to teach topic understanding and thinking simultaneously (McGuinness, 2007). McGuinness (2007) added that infusion has two forms: either it can be specific to any subject (for example science, mathematics, history) or it can be developed to cover more subjects in the curriculum. Infusion across the curriculum is considered as a good strategy by some scholars regarding the benefits it gives for young learners "who can recognise and use common patterns of thinking, deepen their understanding of curriculum topics, make connections between them, and thus be a position to capitalise on new learning opportunities" (McGuinness, 2007; Bruer, 1993).

Beyond enrichment and infusion, we can distinguish the training programs according to its context and content. Some of the researchers believe that reasoning skills must be and can only be taught explicitly (see e.g. Feuerstein, Rand, Hoffman, & Miller, 1980; Klauer, 1989, 1991, 1993; Lipman, 1985), while others believe that operations should be embedded in traditional school subjects (e.g. the CASE project or see Dienes, 1963, 1973; Shayer & Adey, 1981). In the present study, we combine these two approaches by delivering explicit training of inductive reasoning strategies in mathematics content.

2.2.2 Klauer's definition, model, and training of inductive reasoning

Klauer's as well as his colleagues' perspective and theory to inductive reasoning is considered as a well-structured and probably the most detailed model of inductive reasoning (Klauer & Phye, 1994; Klauer et al., 2002). According to his definition "inductive reasoning consists of detecting regularities and irregularities by finding out:

A: {a1: similarity; a2: difference; a3: similarity and difference} of
B: {b1: attributes; b2: relations} with
C: {c1: verbal; c2: pictorial; c3: geometrical; c4: numerical; c5: other} material"
(Klauer & Phye, 2008, p. 87).

Klauer's definition "enables one to detect regularities, rules, or generalizations and, conversely, to detect irregularities. This is one way in which we structure our world" (Klauer, Willmes, & Phye, 2002, p. 1). He distinguished regularities, irregularities, and diversities, that is, if a rule did not cover the whole set of elements, it should be ignored in favor of a more suitable one. A total of thirty cases (3 x 2 x 5) could be formulated based on the above facets. The first facet A is referring to the type of the comparison: similarities, differences or the combination of both, the second facet defines the category of the comparison (attributes or relations among the elements), finally, facet C defines the type of the material.

Table 2.1. Klauer's taxonomy of the classes of inductive reasoning (Klauer & Phye, 2008, p. 88)

Process	Facet identify-cation	Item formats	Cognitive operation
Generalization	a1b1	Class formation	Similarity of attributes
		Class expansion	
		Finding common attributes	
Discrimination	a2b1	Identifying disturbing items	Discrimination of attributes (concept differentiation)
Cross-classification	a3b1	4-fold scheme	Similarity and difference in attributes
		6-fold scheme	
		9-fold scheme	
Recognizing relationships	a1b2	Series completion	Similarity of relationships
		Ordered series	
		Analogy	
Differentiating relationships	a2b2	Disturbed series	Differences in relationships
System construction	a3b2	Matrices	Similarity and difference in relationships

The definition shows the strategy that one can follow to reason inductively about a given problem by scrutinizing the attributes of the objects or the relations between them.

The central facets, the central parts of this definition are facets A and B. The combination of these two facets gives six classes of inductive reasoning processes (see Table 2.1.)

Figure 2.3 visualizes the relationships among the six different strategies of inductive reasoning processes. It shows two separate branches, which are divided in another two branches. If both similarities and differences are called for the two branches come together again. It results in a symmetrical figure because the attributes and the relationships are similarly differentiated in both branches. This definition gives the possibility to design an analytic strategy to be able to solve any kind of inductive reasoning problem (Klauer & Phye, 2008).

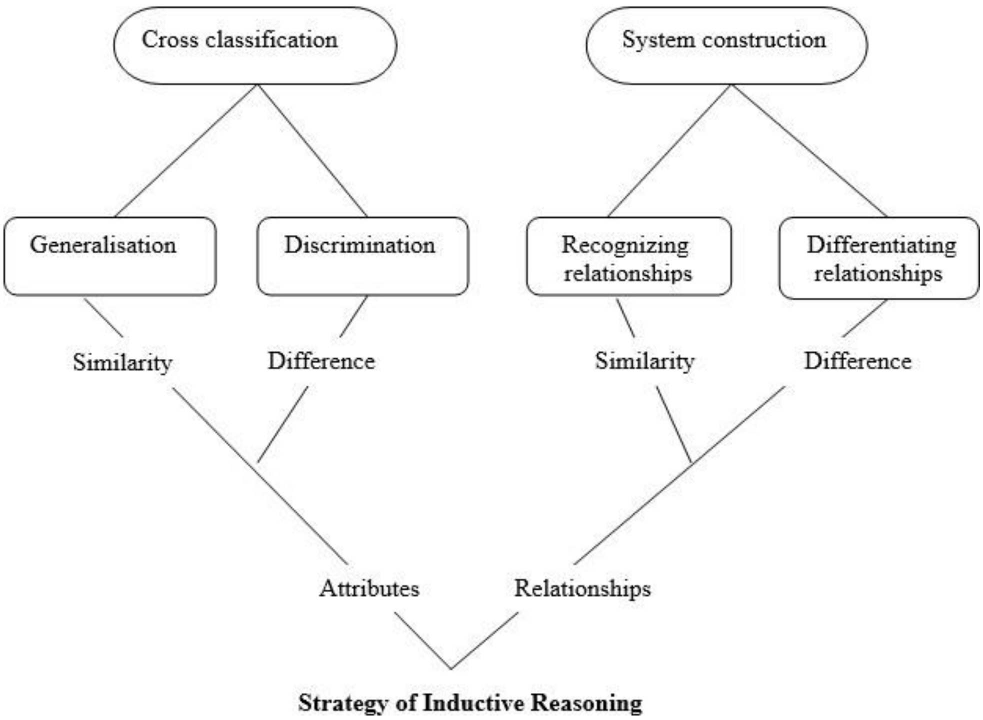


Figure 2.3. Strategies of inductive reasoning (Klauer & Phye, 2008, p. 87)

Computers can be programmed to solve inductive reasoning problems along the same algorithm. Humans are different and may prefer to use heuristic strategies by solving problems of inductive reasoning. Klauer and Phye (2008) depicted the flow diagram of this heuristic strategy usage (Figure 2.4.). According to their flow chart at the beginning of the problem-solving process the participant starts with a global inspection of the problem and builds a hypothesis about the possible solutions. In the next step the previously build hypothesis is tested so that the solution of the problem can be found rapidly. They suggested that in a training program, pupils might be advised to use this heuristic strategy before applying the analytic strategy. The application of the analytic strategy is recommended if the heuristic strategy does not work, does not lead to the right solution. In this process the quality of the tested hypothesis determines the speed of finding the right solution.

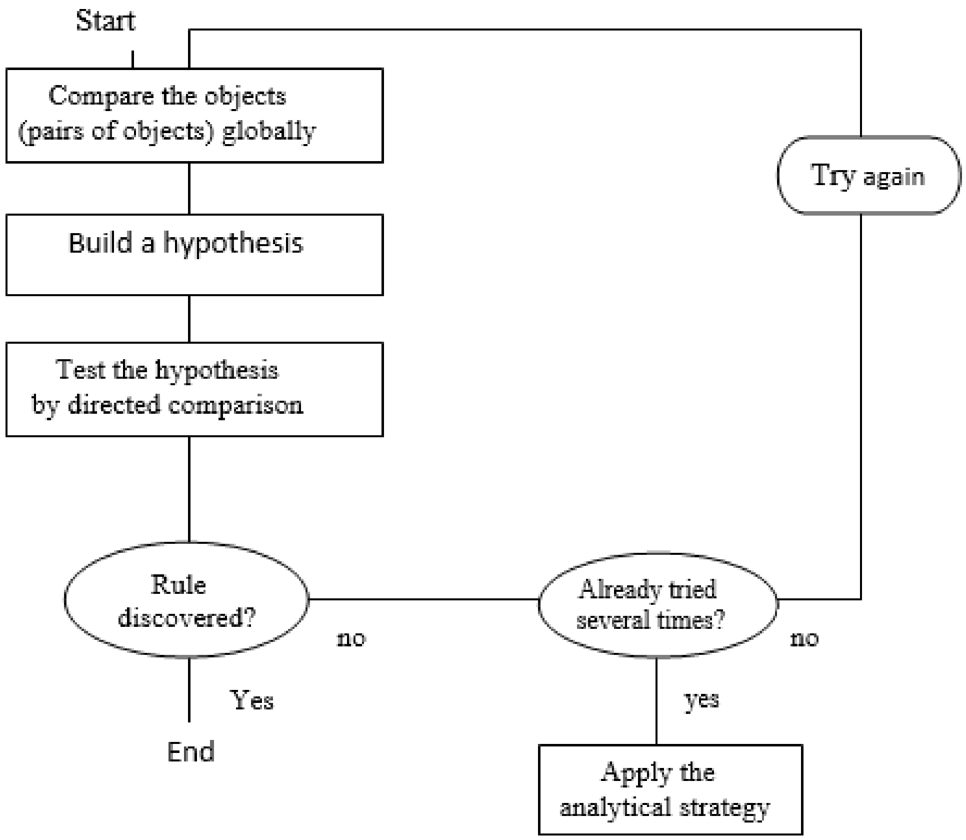


Figure 2.4. Heuristic or hypothesis-guided strategy of inductive reasoning (Klauer & Phye, 2008, p. 89)

To sum up, inductive reasoning is a helpful procedure that is useful in making predictions about new hypotheses. “Inductive reasoning involves making predictions about novel situations based on existing knowledge” (Hayes et al., 2010, p. 278). Therefore, inductive reasoning can be used as a mechanism of hypothesis generating and hypothesis testing (Gilhooly, 1982), and it can also be seen as a means of concept development (Egan & Greeno, 1974; Gelman & Markman, 1987; Markman, 1989).

Klauer developed three cognitive training programmes for pupils: Cognitive training for children I (Klauer, 1989), Cognitive training for children II (Klauer, 1991) and Cognitive training for young children (Klauer, 1993) which is based on his theory of inductive reasoning. The main aim of his work was to teach how to identify the similarities between attributes and relations, and to present the way to solve identified problems by inductive reasoning.

Klauer introduced his trainings to provide the participants a chance to acquire the basic strategies of inductive reasoning, to internalise the strategies by gaining an opportunity to practise those, thus the participants learn how to recognise and solve inductive problems easily (Klauer & Phye, 2008).

All the three training programs (Klauer, 1989, 1991, 1993) are focusing on the six basic classes of IR. The training tasks are divided into groups, each group consists of twenty items for each classes of inductive reasoning. The programs consist of 120 items each. The second and the third level programs include three different problems: verbal, figural, and numerical, each one of them includes forty items. They have been derived from the daily life experiences of the children that also might be in their schools. The training program developed for the youngest pupils is different as it cannot be expected that young children can read the instructions of the tasks.

The effect-size of the programs are presented in Klauer and Phye’s (2008) paper. The trainings were applied in 74 training experiments with over 3600 participants from different age groups. According to the results the programmes had a reasonable transfer effect on fluid intelligence and various academic subjects (Klauer & Phye, 2008).

2.2.3 The adaptations of Klauer’s model for enhancing inductive reasoning skills

The processes and mechanism of IR can be described with a well-structured system and model from an educational perspective (Klauer, 1989, 1990; 1996; Klauer & Phye, 1994; Klauer, Willmes, & Phye, 2002). Pupil’s IR level can be modified very effectively; that is, IR can be trained using tasks developed with this approach (Klauer, 1996; Tomic, 1995; Tomic & Klauer, 1996). A number of studies (Barkl, Porter, & Ginns, 2012; de Koning & Hamers, 1999; de Koning, Hamers, Sijtsma, & Vermeer, 2002; Hamers, de Koning, & Sijtsma, 1998; Klauer & Phye, 2008; Klauer, Willmes, & Phye, 2002; Molnár, 2011; Tomic & Kingma, 1998) have empirically confirmed this statement, reporting a significant improvement of inductive reasoning skills as a result of IR training programmes developed according to Klauer’s model. While their training programmes used different contexts, particular cultures or ages, or different target groups (pupils with special needs and average or gifted pupils), they were each implemented as a face-to-face intervention study.

Bottino et al. (2007) designed a project for second to four grades pupils (age 7-10) to foster their reasoning skills by engaging them in different computer games. The children who participated in the experiment had very little knowledge about computer games and they did not have personal computers. The experiment took place at the computer laboratory during class hours in the pupils' school. Each meeting took about one hour per week and the participating pupils were divided into three groups (high, medium, and low achievers) regarding their achievement at their schools. Several computer games (like Mastermind, Minefield, Battleship, Chinese Checkers, Labyrinths, etc.) were used by the researchers in their study focusing on games that support the development of pupils' thinking by for example requiring to create reasoning strategies to solve a specific problem. The training effect was positive; it developed pupils' logical and strategic reasoning. This type of enhancement, that is, using existing software programs for developing pupils' knowledge and skills helped to reinforce the relations between the teacher and his or her pupils.

Another training program which focused on fostering first graders' inductive reasoning, is Molnár's (2011) training program. The training tasks were based on Klauer's definition of inductive reasoning (the German cognitive training for children (Klauer & Phye, 2008). The sample of the experimental study consisted of 90 pupils for the experimental group and 162 pupils for the control group, all the pupils were first graders. The training consisted of 120 problems as Klauer's original program divided into 20 problems for each class of inductive reasoning. By solving the problems pupils needed to apply appropriate inductive reasoning processes. Because of the young age of the targeted group the problem developers payed a strong attention to the used pictures and objects, that is, the used objects (building blocks, colorful pencils etc.) and pictures matched the targeted group's age. The results showed significant improvement on all the six basic structures of inductive reasoning. Regarding gender, there were no relations between the effectiveness of the program and the gender of the pupils, that is, the training proved to be gender-free.

Pásztor (2014) continued Molnár's work and developed a training program for higher graders to foster their inductive reasoning skills in mathematical content by means of technology. His training program was also based on Klauer's model of inductive reasoning and on his concept of cognitive training for children (Klauer, 1989). Following the original model, the online training contained 120 computerized learning tasks. These tasks were embedded in various mathematical content like recognizing and discriminating relations or attributes through mathematical operations, number series and units of measurements. The effectiveness of the training was tested in an experimental study with third and fourth grade pupils (N=240). The training lasted five weeks. The results of the training program were positive in general. The inductive reasoning skills of the participating pupils in the experimental group increased significantly.

2.3 The effect of the developmental level of inductive reasoning skills on the development of other domains

At the beginning of this millennium, almost the same ideas emerged, but in a new wave of teaching 21st-century skills (Molnár & Csapó, 2019a; Voogt et al., 2013; Wegerif, 2006). The fostering of thinking skills, a set of cognitive skills essential for learning and creating new knowledge, became interesting in the context of educational policy (Wegerif, Li, & Kaufman, 2017; Vainikainen et al., 2015).

Several researches have highlighted at the importance of inductive reasoning in learning processes (Molnár et al., 2013). As inductive reasoning is a general thinking skill, it is closely related to intelligence (Klauer et al., 2002; Klauer & Phye, 2008) and to almost all higher-order cognitive processes (Csapó, 1997), such as knowledge acquisition and application (Bisanz, Bisanz, & Korpan, 1994; Hamers, De Koning, & Sijtsma, 2000), problem solving (Klauer, 1996; Tomic, 1995; Klauer, 1989), hypothesis generation, and hypothesis testing (Gilhooly, 1982), and analogical reasoning (Goswami, 1991; Molnár et al., 2013).

Tests developed for measuring inductive reasoning skills contain mostly analogies, series, classifications and matrices (Goldman & Pellegrino, 1984; Sternberg & Gardner, 1983; van de Vijver, 1991). Klauer et al. (2001). It is the case with intelligence tests too, that is, in most of the cases the working of inductive reasoning processes are needed by solving intelligence test's tasks.

Contemporary cognitive research reveals that the developmental level of inductive reasoning has a significant effect on the success of knowledge acquisition and application

(Bisanz et al., 1994; Goldman & Pellegrino, 1982; Pellegrino & Glaser, 1982; Klauer, 1990; Klauer, 1996; Hamers et al., 2000); therefore, inductive reasoning "is essential for gaining a deeper understanding of any subject matter and its application in real-life problem situations" (Molnár et al., 2013, p. 37). Csapó and Nikolov (2009) examined the cognitive contribution of inductive reasoning to second language proficiency. They found significant contribution of cognitive skills to school learning, especially to the proficiency level in foreign language. According to Cheng and Holyoak (1985) reasoning skills develop the expertise level of the pupils by developing the way, pupils select and apply logical rules in the tasks.

The implementation of inductive reasoning in school curriculum has several advantages: "inductive reasoning is one of the mental tools that is used not only to acquire new knowledge, but also to make the acquired knowledge more readily applicable in new contexts" (Csapó, 1997, p. 612). School learning is expected to be more meaningful, pupils' knowledge is also expected to be more deeply understood and easier to be applied in new contexts and situations.

Inductive reasoning was interpreted by many scholars as a contrast to deductive reasoning. As a result, they launched researches for contrasting inductive and deductive reasoning or comparing them to each other (see Ennis, 1987; Johnson-Laird, 1988; Shye, 1988; Sternberg, 1986).

Inductive reasoning is a component of critical thinking (Ennis, 1987). It is a skill that accommodates deduction, computing, and associations (see Johnson-Laird, 1988). Inductive reasoning is a component skill of problem solving (Wu & Molnár, 2018a). It has a very essential role in almost all steps in information processing: in the selection and application of the proper problem-solving strategy, and in some activities in the decision-making process.

Molnár (2011) suggested that the explicit and/or implicit training of inductive reasoning in the classroom has a very important role in acquiring deeper understanding of the subject matter. Findings suggested that its fostering is not an integral part of the school curricula (de Koning, 2000). Research on contemporary cognition has shown that our knowledge in general is either content-bound or domain-specific, that means the transfer of the knowledge from one filed to another one is not easy (see Csapó, 1997). Csapó (1997) indicated that inductive reasoning itself is a mean of transfer which requires applying the knowledge in one context in a new situation. Experiments focusing on the transfer mechanisms (Klauer, 1989; Phye, 1989) and the level of inductive reasoning in general (Klauer, 1990) confirmed the research results that inductive reasoning can be developed and improved and as a result a higher level learning outcome can be expected (Csapó, 1997). Other researchers emphasized also that inductive reasoning can be developed effectively (Klauer & Phye, 2008; Molnár, 2011). To sum up, the including of reasoning skill activities must be part of the school curriculum (Molnár, 2011; de Konig, 2000; Resnick, 1987).

Studies focusing on the developmental curve of inductive reasoning suggested that the most sensitive period falls to early school years. According to Molnár et al. (2013) and Molnár and Csapó (2011) the development occurs during several years (see Figure 2.5) which gives a wide opportunity to develop and foster these skills in classroom environment. Hotulainen et al. (2016) completed this research finding and highlighted that the development is also noticeable by pupils with special needs, that is, even very low achieving pupils' ability level can possibly be developed to a certain level.

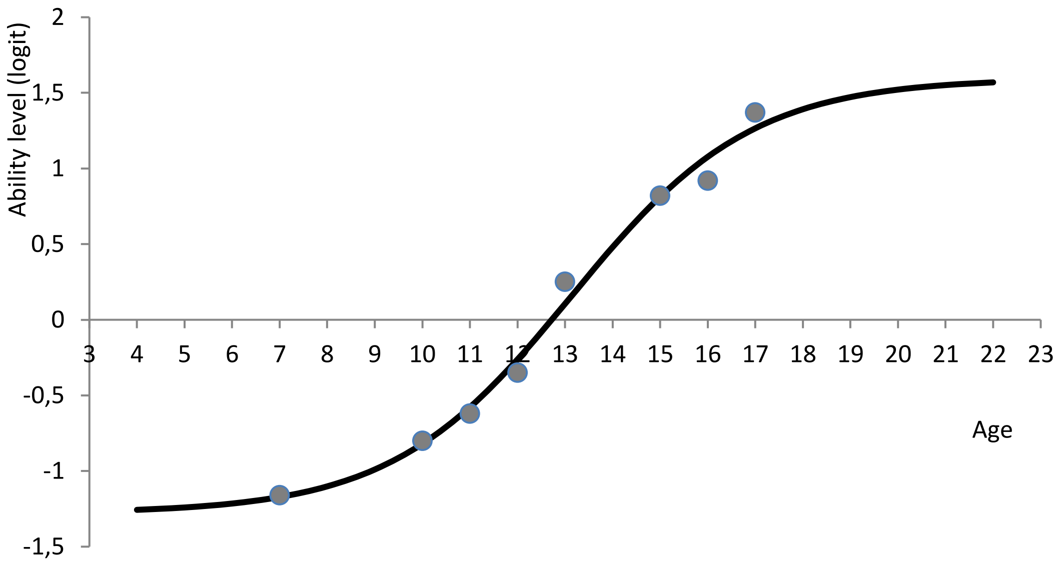


Figure 2.5. The development of inductive reasoning skills (source: Molnár & Csapó, 2011, p.134)

Vo and Csapó (2020) in their study on the development of inductive reasoning among pupils between the ages of 10-16 years found out that the fastest development occurred in case of 12-14 years old pupils and the development decreased as the age of the pupils increased. That means the younger the children are, the bigger the opportunity is to make effective development in IR skills.

2.4 Mathematics and thinking skills

Mathematics is one of the most important subjects which is connected to all sciences. It belongs to the subjects that are taught from the beginning till the end of compulsory education, in elementary, primary and secondary school years – in the majority of education systems worldwide. Therefore, studies on mathematics education are widespread (Adler, 2017; Yesildere & Turnuklu, 2007; Arslan & Yildiz, 2010; Gibney, 2014; Herlina, 2015; Saragih & Napitupulu, 2015; Hudson, Henderson, & Hudson, 2015; Ramaley, 2007). The researchers emphasize the importance to develop mathematical thinking in education since mathematics is an important branch of science that involves developing thinking (Onal, Inan, & Bozkurt, 2017). According to studies, thinking skills can be developed in two different ways (see Chapter 2.3.): separately (Lipman, 1985) or embedded into school subjects (Swartz, 2001; McGuinness et al., 2003; Rajendran, 2010). Mathematics is one of the school subjects that can develop and enhance thinking skills (Rajendran, 2010; Aizikovitsh & Amit, 2010). In addition, mathematical thinking skill is considered as a cognitive skill (Onal, Inan, & Bozkurt, 2017) since both skills are connected to each other. Onal et al. (2017) explained in their study on mathematical thinking skills that the weaker the cognitive skill abilities, the weaker the mathematical skills are. They reached this result when they noticed that the failure of some participants in the mathematical thinking skills in their study was related to the weakness of cognitive skills of the participants.

Pupils need to have the minimum mathematical skills to be able to handle any mathematical tasks. The general rule says, the earlier the better, this can be applied to the development of numeracy for pupils. The earlier the child learns numeracy, the advanced the child becomes in mathematics as studies emphasize the importance of the development of early numeracy and the success that can be achieved afterwards with mathematics knowledge (Pásztor, Molnár, & Csapó, 2015). Early numeracy consists of several basic skills and concepts (Jordan, Kaplan, Locuniak, & Ramineni, 2007). Early mathematical skills have important skills that can be considered as the number word sequence skills (Rausch & Pásztor, 2017). Enumeration skills can be developed by the knowledge of the correct order of number words either forward or backward, it is also related to other things like solving basic additions and subtractions (Aunio & Rasanen, 2015). Many international assessment studies use mathematical content as a main component in the process of their data collection such as the assessment programmes PISA (OECD, 2013) or TIMSS (Mullis & Martin, 2013), it is also used in the national assessment systems around the globe (Pásztor, Molnár, & Csapó, 2015).

The connected skills and abilities to mathematics has been studied carefully for about two decades ago focusing on the assessment or the development (Csapó & Csíkos, 2011). Csapó and Csíkos (2011) added that these skills and abilities are necessary to access to mathematics, for example, a pupil should have a proper level of reading comprehension to be able to go into mathematics, at the same time, the skills of reading comprehension of texts is improved by learning mathematics since the logic between both of them (mathematics and language) influences and improves each other. Some of these studies focused on the relations between mathematical skills and the ability of intelligence and the findings concluded that mathematical achievements are significantly connected to several components of fluid intelligence like general sequential reasoning, quantitative reasoning and the Piagetian reasoning (Carroll, 1993).

It is commonly assumed that the development of reasoning skills is embedded in ordinary school material (de Konig, 2000), which focuses mainly on reading, writing, and mathematics (Molnár, 2011). However, according to Molnár and Csapó (2019a), no appreciable development could be noticed in pupils' reasoning skills in reading and science between grades 2 and 6, and "there was a steady increase detectable in mathematics, especially in the first four years of schooling…. Overall, these results highlighted the importance, sensitivity and potential of the development of thinking skills in the early years of schooling". Therefore, enhancing thinking skills should become a real goal in education (Vainikainen et al., 2015).

According to Csapó and Csíkos (2011) linguistic development is highly correlated with mathematical achievement since people tend to count in a particular language. Counting skills are influenced by the designation of numbers in various languages. They added that mathematical word problems can be understood by oral and written comprehension skills, they also stated that reasoning skills and mathematical thinking are highly relevant. Thinking skills and mathematics are closely connected to each other and the skills can be assessed and developed using mathematical contents by many diagnostic assessment programs which have used thinking abilities in that context including inductive (Csapó, 2002). Molnár and Csapó (2019c) supported the previous point by saying that generic objects and domain specific objects are there in mathematics. They added that the generic covers operational reasoning like seriation, class inclusion, classification, combinatorial reasoning, probabilistic reasoning, proportional reasoning as well as some higher-order thinking skills like inductive reasoning and problem solving, and all of them can be assessed by mathematics. They added in literature, that there are very few studies that focus on the cognitive processes like reasoning skills and its connection with understanding mathematics, and the main focus of these studies was on the most hidden aspects of learning.

In their study on the psychological dimension of learning (cognitive development and reasoning), Molnár and Csapó (2019c) reported thinking and reasoning can be assessed separately from the other dimensions (application and disciplinary knowledge) (see Csapó & Szendrei, 2011; Csapó & Csépe, 2012; Csapó & Szabó, 2012) with a good validity. They also noted that the psychological dimension of learning can be measured in early years in the context of mathematics as one of the most important domains of learning. They added that the cognitive development of pupils can be effectively enhanced if it is taught intensively (see also de Koning et al., 2002; Klauer & Phye, 2008; Perret, 2015) and that enhancement can be seen mainly in the first years of schooling, in the domain of mathematics. In other words, the findings indicate that the development of the psychological dimension of learning can be achieved in school knowledge in mathematics.

Pásztor (2014) presented in his study on fostering inductive reasoning through mathematical content using technology environment that integrating mathematical content and reasoning strategies through learning tasks is possible even in computer-based settings. The findings of the study showed that even with the current form of the training program in mathematical content, the development of the pupils' IR thinking skills are possible and achievable at early age and for all types of pupils regarding their achievement, whether they are weak, average or high achievers.

2.5 Influential factors to students' achievement

The effect of background variables e.g. mothers' educational or occupational backgrounds, socio-economic status (SES) etc. might differ from one country to another (see Csapó, 2010; Nikolov & Csapó, 2018; Pásztor, 2016). There are many factors that might play a role in this change, such as culture and environment, therefore, different indicators are used to categorize these factors. Background factors sometimes impose serious changes to the results so some international studies like PISA took it in consideration in their data collection and used some of the factors. PISA used factors like social, cultural, and economic in their studies (OECD, 2016). In addition, the variables differ in influence as well. A specific variable in a specific region might have high or low influence on the performance of a group of people (e.g. school children). In other words, the nature of a specific place makes a variable influential and vice versa (see Csapó, 2010). Csapó (2010) explained in his study that the mothers' educational background is the most significant on pupils' performance among the other variables in villages and cities in Hungary. There are differences between the two types of settlements (villages and cities) but that did not have any influence on pupils' achievement, even though the type of settlement is different and the achievement of the pupils is different comparing those who live in villages and the others who live in the cities (Nikolov & Csapó, 2018). Therefore, the best to explain that difference was the parents' educational level (see also Nikolov & Csapó, 2009). Pásztor (2016), in the results of the training program for the development of inductive reasoning thinking skill found out that the achievements of the participant pupils were strongly influenced by two main things: parents' education and pupils' socioeconomic background.

Another important factor that might play a role in changing the results is gender (see Weaver & Raptis, 2001; Clariana & Wallace, 2002; Hotulainen, et al., 2016). Clariana and Wallace (2002) in their study did not find any relation between the gender of the participating pupils and the performance difference. Weaver and Raptis (2001) in their study supported the results of the previous study by stating that in their test they did not have any significant effect difference on the performance of males or females and that was the case for all the components of the test. Molnár (2011) in her training program in inductive reasoning for first grade pupils found out that there were no significant differences between boys and girls in any domains of the measurement points she used in the study. Pásztor (2016) in his training program for assessing and developing inductive reasoning by the mean of technology found out in the analysis of the results that the gender of the participating pupils have not caused significant differences meaning that being male or female will not change anything in the achievement. He got the same results with no gender differences in another study on fostering inductive reasoning through mathematical content using computer-based environment.

Some other studies emphasise that gender might play a role in achievement differences. Molnár and Csapó (2019b) regarding identifying early mathematical learning difficulties showed a significant gender-level achievement differences in favour of the girls. They said that the difference change occurred between the different dimensions. Csapó, Lőrincz, and Molnár (2012) in their study on the innovative assessment technologies in educational games for young pupils found in general no differences in the performance while when it came for deeper analyses, a significant difference was found in a specific group of pupils. Therefore, it can be concluded that gender as a background factor might have influence on the achievement level of the pupils.

Other frequently analysed influential factor is the age/year of pupils (see Pásztor, 2016, Csapó, Molnár, & Kinyó, 2008; Molnár, Greiff, & Csapó, 2013) in the development and correlation of thinking skills and the acquisition level of the school curriculum. Or the developmental level of inductive reasoning and its influential effect on the efficacy of the teaching and/or training programs. In other words, if the children were classified to, for example, three levels (low, medium, and high) regarding their achievement (see Pásztor, 2014) in their schools or their achievement in a pretest (if it was a training program), the researchers are able to get more results regarding each group and that provides deeper analyses to which groups are the most affected by the training. The results concerning such a factor can only be generalised in a specific region where the study took place since all pupils have the same characteristics i.e. the same curriculum. Thus, the results cannot be generalised to another environment since the children there have totally different characteristics i.e. curriculum, tasks, teaching methods and so on (see Mayer, 1992; Funke & Frensch, 2007).

The importance of socio-economic status can be seen when it explains the reasons in the differences between the performance of the pupils as it was the case in some of the studies mentioned earlier. Thus, knowing the variable that makes the variance in the achievement gives better results with detailed explanations. In Palestine, we do not have a clear data regarding the SES and the pupils' achievement therefore we decided to use several variables (e.g. gender, grade, mothers' educational and occupational backgrounds, school achievement) in our studies and find the most influential ones, that will give us more accurate results.

2.6 Cross-cultural validation

The necessity of a cross-cultural validation appears when adapting instruments for use in a culture different from what the original ones have been developed for. Ensuring whether the originally developed instruments in a particular culture are meaningfully applicable and equivalent to be used in a totally different culture is called cross-cultural validation (Huang & Wong, 2014). Therefore, any instruments aimed to be used in another culture that the one it has been created for should be adapted in a certain way to be applicable for the new culture it is intended to be used in.

In psychological studies, cross-cultural validation has been used as a generally used method. In several times it means a simple translation of the instructions to the target language. In case of other studies that use instruments with visual pictures, drawings, figures, elements which can be culture-dependent, a simple translation of the instruction is not enough. A more thoughtful adaptation of the tasks, instrument is needed (see Wu & Molnár, 2018a; Wu & Molnár, 2018b).

The Arabic context is different from the European context, so cross-cultural validation requires adaptation in several factors (Mousa & Molnár, 2018; Mousa, & Molnár 2019a; Mousa & Molnár 2019b). A simple translation of the instructions (from English to Hungarian – or any European language) is not enough because of the large cultural differences. Arabic is basically a right to left and not a left to right language, which result in direction issues within the tasks (e.g. by placing the stimulus and the answer possibilities). Indian numerals (used in Arabic countries) differ from the Arabic numerals (used in European countries), which also results in extra changes in tasks, using mathematical symbols. To sum up, the elimination and minimization of the factors that might influence the results of any study requires well-deliberated and empirically proven decisions regarding the test adaptation process.

2.7 Summary

The present empirical research is routed in two main research topics: technology-based assessment and enhancement, and inductive reasoning skills. Technology offers and opens new opportunities in instructional and developmental processes, including educational assessment. By now, in most of the large-scale assessments computer-based assessment replaced traditional paper-based testing, expanding the possibilities to include more domains, even 21st century skills, in the assessment process and widening the possibilities educational assessment can offer for the teaching and learning process. Parallel to this issue, game-based learning also became a hot issue in cognitive educational researches. Several researchers demonstrated its active role as a useful learning tool.

Still there is no universal definition for inductive reasoning skill (see Klauer, 1990; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Sloman, 1993; Gick & Holyoak, 1983). However, the most practical and detailed model was developed by Karl Joseph Klauer (1989). It offers a well-deliberate definition classifying all tasks and item types which solution needs the use of inductive reasoning operations. In this regard, the psychological concepts make Klauer’s theory an excellent practical and applicable theory in everyday classroom teaching and it proves to be handy to all school subjects.

The results of the empirical studies in the field of inductive reasoning were positive in general (Barkl, Porter, & Ginns, 2012; de Koning & Hamers, 1999; de Koning, Hamers, Sijtsma, & Vermeer, 2002; Hamers, de Koning, & Sijtsma, 1998; Klauer & Phye, 2008; Klauer, Willmes, & Phye, 2002; Molnár, 2011; Tomic & Kingma, 1998). Development was noticed by the studies on various parts of pupils’ cognitive skills i.e. pupils’ logical and strategic reasoning, as well as fluid intelligence and various academic subjects. Inductive reasoning can be separated and distinguished from other reasoning skills, however its influential power is significant. In the educational context it is a frequently asked research question how the developmental level of inductive reasoning influences pupils’ cognitive skills like fluid intelligence. Furthermore, according to the research results the sensitive period for its development falls to early school years, however, its development spans during the whole compulsory education, offering great possibilities for successful school trainings in a wide range of school grades.

Many experimental studies emphasised the importance of reasoning skills, especially inductive reasoning, in many directions regarding knowledge improvement and learning enhancement (see Goldman & Pellegrino, 1982; Bisanz, Bisanz, & Korpan, 1994; Hamers, De Koning, & Sijtsma, 2000; Klauer, 1990, 1996; Pellegrino & Glaser, 1982). Studies in this field provide evidences on the strong relationship between inductive reasoning and other higher order thinking skills such as critical thinking (Ennis, 1987) and problem solving (Wu & Molnár, 2018a). Research results have also highlighted the possibility to develop inductive reasoning skills, which can have a positive impact on other fields of education, like subject leaning (see Csapó, 1997; Csapó & Nikolov, 2009; Willmes & Phye, 2002; Klauer & Phye, 2008).

Mathematics is probably the only subject, which exists in all education systems and the learning of which spans over the whole schooling. Mathematics and its enhancing effect have been studied in connection with thinking skills. According to the research results, learning at school, learning the different subject materials can improve and improves the level of thinking, however in most of the cases it happens as a by-product of teaching and not as an effect of an explicit training of thinking skills and operations. Learning mathematics seems to have the highest impact on the development of pupils' thinking skills, including inductive reasoning. Studies emphasized mathematics to be one of the best domains to measure pupils' psychological dimension of learning even in early years of schooling.

There are several factors which can influence learning outcomes and pupils' test results. The most widely assessed and analysed factors are the socio-economic background factors, referring to parents' education and some others like the gender.

3 THE STRUCTURE OF THE PALESTINIAN EDUCATION SYSTEM

The Ministry of Education and Higher Education (MoEHE, 2017)[2] is responsible for all regulations relating to public schools in Palestine. It also interferes in the regulatory overview of private and UNRWA[3] schools. The same case is observed with the Palestinian universities and community colleges where MoEHE set the regulatory overview with the higher education sectors.

The Palestinian general education system is divided into three different sub-sectors (see Figure 3.1.):

1. Kindergarten education: Local and international institutions provide services in this sector for children in the age of 4–5 years old.
2. Primary education: It starts from the first grade until the tenth grade. This sub-sector is divided into two levels:
   a. The first level is grades 1–4 called the lower basic stage.
   b. The second level is from grade 5–10 called upper basic stage.
3. Secondary education: It starts from grade 11–12, and it includes two streams: academic and vocational education. Pupils can choose between these two streams. At grade 12, pupils attend the general examination which is based on their final result. Pupils who pass the exam can apply to universities and colleges.

The Palestinian education system is still working on improving the quality of education through the implementation of technology in everyday school practice (Shihab, 2014). Implementing technology at schools in Palestine covers several areas: offering training courses for teachers; connecting all schools to the internet; digitalizing learning materials (e-books) and providing the courses with educational videos and computer games. According to the initiated developments (2017-2020 plan), each teacher and pupil will have their own tablet by 2020 (Alhadath, 2016). It should be kept in mind that some

[2] All information provided under the title "The structure of the Palestinian education system" is collected from the official website of the Ministry of Education and Higher Education. It was referred to in the references (MoEHE, 2017).

[3] UNRWA stands for The United Nations Relief and Works Agency for Palestinian Refugees. It is funded almost entirely by voluntary contributions from the United Nations (UN) member states. It runs schools under the name of UN. For more info, see UNRWA (2017).

schools have already started using tablets, before this 2017-2020 educational development plan 35 thousand tablets were handed out to pupils (Shihab, 2014), because ICT is considered as a powerful tool regarding developing logical abilities (Riel, 1994). Nowadays, it is important to find out what additional values can modern technologies bring to the Palestinian education system.

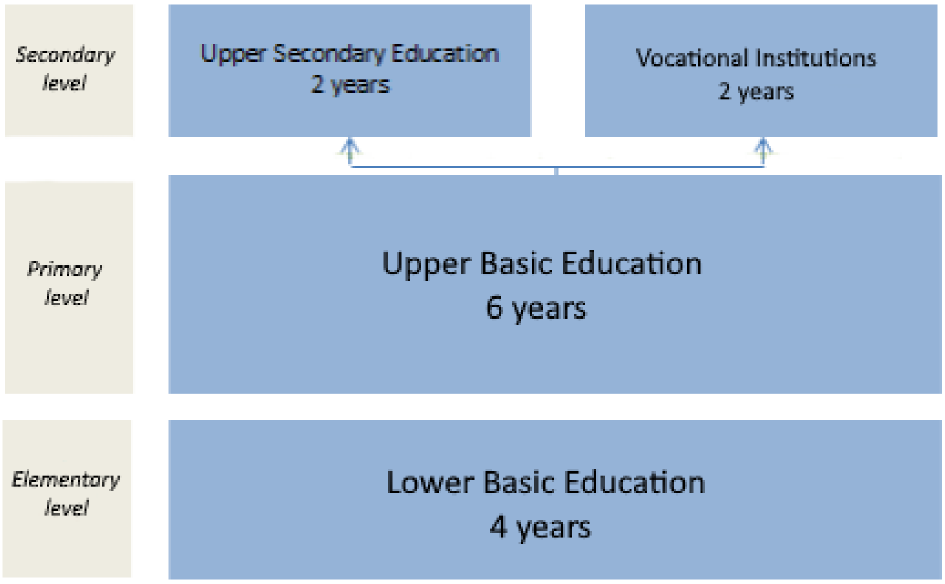


Figure 3.1. The Palestinian school education system (based on World TVET Database Palestine, 2015).

The Palestinian Ministry of Education and Higher Education considered technology as an important factor in education to keep up to date with the new educational models. The implementation of Information and Communication Technologies (ICT) at schools started in 2002 when the ministry of education designed and implemented related initiatives to include technology as much as possible in teaching and learning processes and that includes creating capacity building, infrastructure, e-content, technology curriculum etc. (Shraim, 2018). Statistics show that around eighty percent of the governmental schools are well equipped with computer laboratories and more than two-thirds of these labs are connected to the internet (PMEHE, 2018). However, information and communication technologies in the Palestinian education system (with its importance to overcoming restrictions on movement, sharing knowledge, and combating unemployment) still needs a lot of improvement and enhancement to be visible (see Pacetti, 2008).

Subsequently, there are some studies which analysed the Palestinian curriculum regarding higher order thinking skills. These studies emphasised the visibility of these skills in the curriculum on an average level (AbdulKader, 2014). There are also other studies that explore the educational curriculum and the enhancement of thinking skills. The findings highlight at weaknesses in the curriculum and in teaching and assessing pupils' thinking skills enhancement (Barbak, 2012).

4 AIMS AND STRUCTURE OF THE EMPIRICAL STUDIES

Several countries around the World are developing their education systems by involving technology in it. Thus, Palestine is seeking to be one of these countries by establishing a digitized education system (Affouneh, 2014). Up to the present time, there is no evidence or overview of the development of higher order thinking skills of Palestinian school pupils. Therefore, this study's main objective is to assess and enhance Palestinian school pupils' thinking skills (mainly inductive reasoning) in their early school age by the means of technology. We also wished to explore the possibility of applying online computerised tests that already have established psychometric characteristics to assess pupils in regular Palestinian educational practice at the early stages of schooling. We started to explore the feasibility and the applicability of computer-based assessment among young pupils by testing their basic mouse skills then moving forward to test their thinking skills in order to find out if there are any possibilities to apply enhancement training programs. This study connects to important developing areas of educational research and it places them in the context of the development of the Palestinian education system. First, improving the quality of thinking skills in the Palestinian educational context, especially when it comes to early-age school-children and that can be fulfilled by paying attention to these skills and the importance of enhancing pupils' performance. Second, more attention is being given to educational assessment in research and practice, so measuring certain psychological constructs opens the doors to do further researches and training experiments on a specific field benefiting from the feedback provided by the assessment which might help directing some educational practices. Third, computer-based testing makes more constructs measurable, it reduces the timeframe of the assessment and makes it less costly.

The dissertation consists of four empirical studies, building on one another. The first study is about the feasibility of computer-based testing in Palestine among lower primary school pupils: assessing mouse skills and inductive reasoning. The aims of this study were to introduce and explore the possibility of using computer-based testing in Palestinian schools. It investigated the developmental level of mouse skills among year 2 and 3 pupils and tested the applicability of an online test measuring pupils' inductive reasoning. For this purpose, we applied online tests via the eDia assessment platform that already has established psychometric characteristics. We adapted the tasks of both tests to Arabic in both sense: language and directions and monitored whether computer-based assessment is applicable at the Palestinian context or not.

The second study explored the possibilities of applying computer-based assessment in regular Palestinian educational practice at early school ages. We wanted to validate the results of the previous study. So, we applied the same test of inductive reasoning on a larger sample and extended the age-range of the sample to 4th graders.

The third study investigated the possibility of applying computer-based assessment in the Palestinian educational context, in the Palestinian schools among 4th and 5th grade pupils. A Hungarian online test adapted to Arabic format in language and direction has been applied to test and detect the applicability of computer-based assessment in the Palestinian schools. This test was developed from the previous study, more items were added to this test including numerical analogies and numerical series items. Beyond applicability, the present paper addressed the relationship that inductive reasoning has with academic achievement. For this study, we developed the online test further, including more figural and numerical items.

The fourth study examined the applicability and effectiveness of an online inductive reasoning training programme in maths, based on Klauer's "Cognitive training for children" concept in the 9–11 age range and in an Arabic educational context. We investigated the effectiveness of our intervention programme on different groups of pupils, that is, on pupils with different starting levels of IR, on pupils with different socio-economic factors, and on pupils of different genders. We have used a Hungarian online test and training after adapting both to Arabic format in language and direction to find out the effectiveness of the computer-based training in the schools in Palestine.

To sum up, the aim of the studies were to detect the applicability of computer-based testing in Palestine even among lower graders and to develop an inductive reasoning test with a training program, which can be used in everyday school education for enhancing Palestinian pupils' inductive reasoning skills. Table 4.1 summarises the timeline of the empirical studies.

Table 4.1. The timeline of the empirical studies

Timeline	Research activities	Instruments	Samples
February 2017	Piloting the assessment instruments in the schools Collecting the related literature	Mouse skills (Molnár & Pásztor, 2015b) Inductive reasoning items (Molnár et al., 2013; Pásztor et al., 2017; Molnár & Pásztor, 2015b) Background questionnaire eDia online system	2nd & 3rd grades *N* = 57
September 2017	Piloting the assessment instruments in the schools again with a bigger sample size to validate the results of the first pilot study.	Inductive reasoning items (Molnár et al., 2013; Pásztor et al., 2017; Molnár & Pásztor, 2015b) Background questionnaire eDia online system	2nd, 3rd & 4th grades *N* =193
August & September 2018	Piloting the enhancement program and testing the pupils' pre and post to the training	The online data collection was carried out via the eDia platform (Pásztor, 2016; Molnár, Pásztor, & Csapó, 2019; Csapó & Molnár, 2019; Molnár & Csapó, 2019a)	4th & 5th grades *N* = 236

4.1 Research questions and hypotheses

In the followings the research questions and hypotheses of the four studies are listed in the following order:

4.1.1 Research questions for Study 1 (RQ1-RQ7)

1) What are the psychometric properties of the inductive reasoning test? Will the test be reliable?
2) Are pupils able to perform the basic mouse skills in their early age?
3) Is computer-based assessment feasible and applicable among young learner in the Palestinian schools?
4) Which background factors influence the applicability of computer-based testing among 2nd and 3rd year Palestinian pupils?

5) Are there gender differences regarding pupils' mouse and keyboarding skills and their achievement on the inductive reasoning test?
6) How well do grade two and three pupils perform on the inductive reasoning test?
7) Is there any relationship between pupils' school achievement and the developmental level of IR skills?

4.1.2 Hypotheses for Study 1 (H1-H7)

H_1. We expect the psychometric properties of the tests to be acceptable in general.

H_2. Pupils are expected to be able to perform the basic mouse skills effectively (see Clariana & Wallace, 2002) even at a young early age.

H_3. Computer-based assessment is expected to be applicable in the Palestinian schools and we also expect that among young learners.

H_4. It is expected that the background variables related to the mother will have a significant influence on the pupils' performance (see Csapó, 2010; Nikolov & Csapó, 2009; Pásztor, 2016) i.e. more educated mothers, better performance in the IR test.

H_5. We expect to find no gender differences between boys and girls at that early age in their mouse and keyboarding skills or in their achievement on the inductive reasoning test (see Weaver & Raptis, 2001; Clariana & Wallace, 2002; Molnár, 2011; Pásztor, 2016; Csapó, Lőrincz & Molnár, 2012).

H_6. We expect 3rd graders to perform significantly better than the 2nd graders.

H_7. We expect to find a strong relationship between the school achievement of the pupils' and the developmental level of their IR skills.

4.1.3 Research questions for Study 2 (RQ8–RQ13)

8) What are the psychometric properties of the IR test? Will the test be reliable?
9) Is computer-based assessment applicable among pupils in their early age?
10) How well do grade two, three and four pupils perform on the inductive reasoning test?

11) How does female pupils' performance differ from male pupils' performance in the inductive reasoning test?
12) How do background variables (mainly the mother's background) influence pupils' performance?
13) Does school achievement influence the developmental level of IR skills?

4.1.4 Hypotheses for Study 2 (H8–H13)

Our hypotheses for study 2 are based mainly on the results we achieved in the first pilot study (Mousa & Molnár, 2019a), as follows:

$H_{8.}$ The psychometric properties of the tests are expected to be acceptable.

$H_{9.}$ We expect computer-based assessment to be applicable in the Palestinian schools for early age pupils (7-9 years old) based on the results we got from the first pilot study.

H_{10}. We expect the 4th grade pupils to perform significantly better than the 3rd, and the 3rd better than the 2nd graders.

$H_{11.}$ No gender differences between boys and girls at this early age are expected in their achievement on the inductive reasoning test.

$H_{12.}$ We hypothesized that pupils whose mothers have higher educational and occupational backgrounds will perform significantly better than pupils whose mothers have lower backgrounds.

$H_{13.}$ We expect pupils who have higher achievements at their schools to perform better in the IR test.

4.1.5 Research questions for Study 3 (RQ14-RQ18)

14) What are the psychometric properties of the IR test? Will the test be reliable?
15) Is computer-based assessment applicable among fourth- and fifth-graders at schools in Palestine?
16) Is there any tangible development in inductive reasoning between fourth- and fifth-graders?
17) Are there any detectable gender differences?

18) Are pupils' school achievement and their level of inductive reasoning skills and the educational level of their mother related, especially, does the level of pupils' inductive reasoning skills predict academic achievement?

4.1.6 Hypotheses for Study 3 (H14-H19)

$H_{14.}$ We expect that the psychometric properties of the tests are good and acceptable (Pásztor & Molnár, 2013).

$H_{15.}$ Based on the results from studies 1 and 2 (Mousa & Molnár, 2019a, 2018), we expect computer-based assessment to be applicable among pupils in grades 4 and 5.

$H_{16.}$ 5th graders are expected to perform better in the inductive reasoning test compared to the 4th graders.

$H_{17.}$ Based on the results from the previous pilot studies, we expect to find no gender differences between boys and girls at this early age in their achievement on the inductive reasoning test.

$H_{18.}$ We expect again that the background variables have a strong influence on pupils' performance in the IR test.

$H_{19.}$ It is also expected that pupils with higher achievements at their schools perform better in the inductive reasoning test.

4.1.7 Research questions for Study 4 (RQ19–RQ23)

The objective of this study is twofold. Firstly, we examine the applicability and effectiveness of an online inductive reasoning training programme based on Klauer's "Cognitive training for children" concept in the 9–11 age range and in an Arabic educational context. Secondly, we investigate the effectiveness of our intervention programme on different groups of pupils, that is, on pupils with different starting levels of IR, on pupils with different socio-economic factors, and on pupils of different genders.

19) How effectively can Klauer's training concept be adapted to an online environment using mathematical task content?

20) How effective is the online training programme in inductive reasoning in an Arabic educational context for the age range of 9–11?
21) Which starting level of IR is most sensitive to the training programme? In other words, which level can we expect the largest effect on?
22) Which group of pupils can benefit the most from the training programme, according to socio-economic backgrounds, school achievement and according to gender?
23) How can the results be generalised? Are the effects confirmed by latent level analyses using a latent change model in the intervention group and a no-change model in the control group?
24) What do participating pupils in the experimental group think of the training?

4.1.8 Hypotheses for Study 4 (H20–H24)

$H_{20.}$ We expect our developed training programme which is based on Klauer's training concept to be effective even in an online environment (see Molnár, 2011) using mathematical task content (Pásztor, 2014).

$H_{21.}$ We expect our online training programme in inductive reasoning in an Arabic educational context for the age range of 9–11 to be effective (see Klauer & Phye, 2008).

$H_{22.}$ The training programme is expected to have influence on the levels of all pupil groups (see Pásztor, 2014; Molnár & Pásztor, 2012; Klauer & Phye, 2008; Molnár, 2006).

$H_{23.}$ Based on the results of the previous studies, we expect the socio-economic backgrounds to have an influence on pupils' achievement and we expect to see no gender differences between females and males in their achievement on the inductive reasoning test (see also Molnár, 2011; Csapó & Nikolov, 2009).

$H_{24.}$ We expect to find no significant variability among pupils in responding to the intervention programme.

$H_{25.}$ We expect pupils to be fully engaged in the game-based training style and to like it (see Yang & Chang, 2013).

5 METHODS OF THE EMPIRICAL STUDIES

This chapter contains a general description about the methods used in the empirical studies. More elaborated descriptions are available by the detailed introductions and analyses of the given studies (in Chapter 6).

5.1 Samples of the studies: a general description

The samples were selected from schools in the directorate of Bethlehem city in Palestine. We tried to achieve the best representation of schools in the area in terms of quality and type of schooling. We wanted to split the participants into three levels regarding their performance so whole school classes were chosen for group testing since classes in the Palestinian schools are reorganised each year by remixing pupils from different sections to have the ideal curve regarding pupils' ability (advanced, intermediate and low) in each class. Based on their school achievement, pupils were divided into three groups: low, intermediate, and high performers.

In the first study the target group consisted of pupils at the very beginning of schooling. In the second and third study we decided to involve higher graders too, finally, the training program was piloted by four- and fifth-graders (see Table 5.1).

Table 5.1. The samples of the studies

Studies	Samples	Instruments
Study 1	Grades = 2 & 3 *N* = 57	Adapted from the Hungarian item bank (Molnár et al., 2013; Pásztor et al., 2017)
Study 2	Grades = 2, 3 & 4 *N* = 193	Adapted from the Hungarian item bank (Molnár et al., 2013; Pásztor et al., 2017; Molnár & Pásztor, 2015b)
Study 3	Grades = 4 & 5 *N* = 248	Inductive reasoning items were administered via the online assessment eDia (Molnár & Csapó, 2019a).
Study 4	Grades = 4 & 5 *N* = 236	The training is based on a Hungarian training IR programme (Pásztor, 2016). The computer-based IR training was administered via the eLea online training platform (Molnár, Pásztor, & Csapó, 2019). The pre-test and post-test were delivered through the eDia online assessment platform (Csapó & Molnár, 2019; Molnár & Csapó, 2019a).

5.2 Instruments of the empirical studies: a general description of the test development

In the first study, we used two figurative tests: a mouse skill test and an inductive reasoning skill test. The tasks were adapted from a Hungarian item bank. Figure 5.1. shows two tasks from the mouse skill test. In the tasks, pupils could listen to the instructions using headphones and the tasks required pupils to do the operations of clicking and drag-and-drop.

Figure 5.1. Sample items from the mouse test (instruction for the left item: students need to click on the apple and the cup. Instruction for the left item: students need to click on the hole in the bucket.)

Figure 5.2. shows two items from the inductive reasoning test applied in the first study. The items have been adapted to be suitable to the Palestinian context in format and language. Similar to the tasks of the mouse skill test, pupils received the instructions in Arabic using the headphones. The tasks were adapted to the Arabic context regarding the used direction in the tasks too (see Figure 6.3 in Chapter 6.1.2.2). The inductive reasoning items were in two formats: series and analogy.

In the second study, the main focus was on extending the age-range of the pupils and measuring their inductive reasoning. We changed the way the instructions were delivered to written texts as presented in Figure 5.3, so a certain level of reading comprehension was also required to understand what is needed to solve the task. We kept the items exclusively on figurative level.

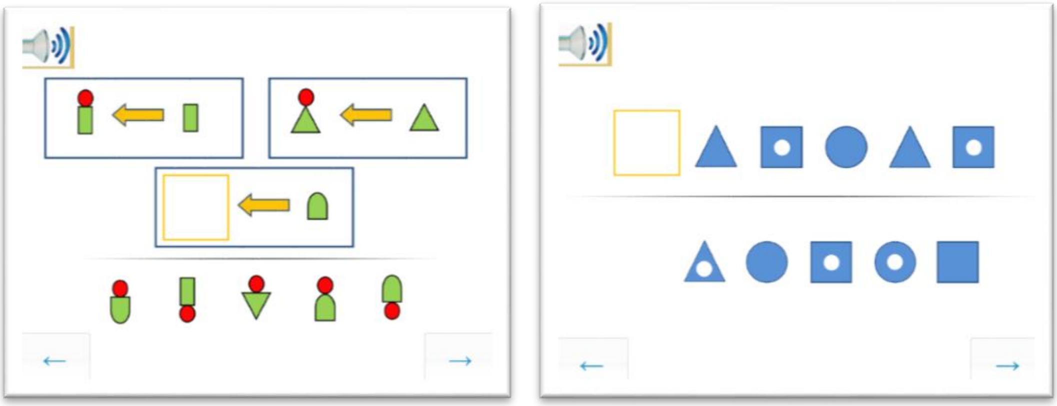


Figure 5.2. Inductive reasoning items from the first study (Instruction for the left item: students need to figure out the rule from the above two boxes to be able to solve the third box. Instruction for the left item: students need to figure out the rule from the series to be able to fill in the empty yellow box.)

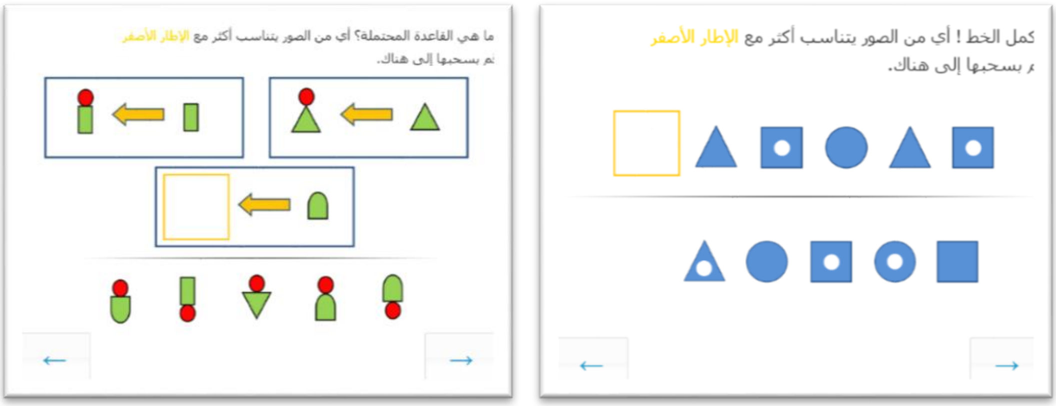


Figure 5.3. Sample items from the inductive reasoning test in the second study (Instruction for the left item: students need to figure out the rule from the above two boxes to be able to solve the third box. Instruction for the left item: students need to figure out the rule from the series to be able to fill in the empty yellow box.)

Based on the analyses, time-on-task, and achievement data, we decided to develop the IR test further by including more and more difficult items too. Apart from the figural items we included number series and number analogy items too. We adapted these items also from the Hungarian item bank developed in the eDia system, based on the psychometrical properties of the tasks (difficulty level and discrimination index). Some adaptation was again necessary regarding the language of the instructions and the numeral system as it can be seen in Figure 5.4. The Eastern Arabic numeral system was applied as it is used in the Palestinian school textbooks in order to eliminate any obstacles that might be related to the shape of the numbers.

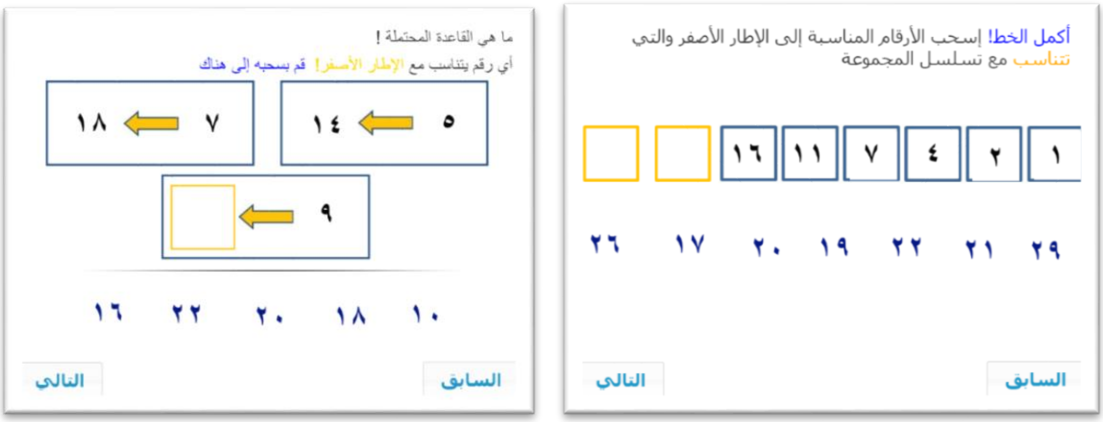


Figure 5.4. Sample items from the number analogies and number series subtest administered in the third study (Instruction for the left item: students need to figure out the rule from the above two boxes to be able to solve the third box. Instruction for the left item: students need to figure out the rule from the series to be able to fill in the empty yellow two boxes.)

Parallel, we started to work on the development of the training tasks. We started to adapt tasks developed on behalf of a Hungarian training programme for inductive reasoning (Pásztor, 2016) and based on these tasks we developed new training tasks too, whose content fit to the Palestinian school curriculum. The structure of the final training programme is similar to Klauer's original training. It consists of 120 learning tasks using the following operations: generalisation, discrimination, cross-classification, recognising relations, discriminating between relations, and system formation. Table 5.2. summarizes the instruments used in the studies.

A background questionnaire (see appendix C) was given to the participating pupils to fill in the requested information with the help of their parents or any family member. The mothers' background was on focus because we expected it to play a major role in influencing pupils' inductive reasoning skills in case of the Palestinian school children. We looked at other factors as well (gender difference and school achievement) for the purpose of comparison.

The instruments in the four studies were completely adapted to the Palestinian context in format and in language as mentioned earlier. Professional teachers in Hungary and in Palestine helped in recording and translating the texts from Hungarian to English and from English to Arabic. Professional programs were also used to change the direction of the tasks without any loss.

Table 5.2. The instruments used in the four studies

Study	Grade	Instrument	Method	Description
Study 1	2 & 3	Mouse usage & IR (Molnár et al., 2013; Pásztor et al., 2017)	eDia platform	Items in the mouse usage tasks require pupils to do clicking and drag-and-drop. IR tasks - domain general and culture-free content reasoning tasks.
Study 2	2, 3 & 4	IR (Molnár et al., 2013; Pásztor et al., 2017; Molnár & Pásztor, 2015b)	eDia platform	IR tasks - domain general and culture-free content reasoning tasks.
Study 3	4 & 5	IR (Molnár & Csapó, 2019a).	eDia platform	IR tasks - domain general and culture-free content reasoning tasks. Numerical items included.
Study 4	4 & 5	IR training (Pásztor, 2016). (Molnár, Pásztor, & Csapó, 2019). IR test (Csapó & Molnár, 2019; Molnár & Csapó, 2019a).	eDia platform & eLea online training platforms	IR tasks - domain general and culture-free content reasoning tasks. Numerical items included. IR training - it consisted of 120 learning tasks as Klauer's original programme.

5.3 Procedures: a general description of the applied procedures

This study depends mainly on collecting data using modern educational technology, which is the current trend in research, therefore, we carried out our research through the eDia platform for all the studies from Study 1 to 4. Pupils did the assessment and the training programme inside their schools in the computer labs which are equipped with modern desktop devices connected to the internet. The training lasted six weeks.

The data collection could not have been performed without taking the official permission (see appendix A) from the directorate of education in Bethlehem. Therefore, we have applied for these permissions by writing an official letter explaining what exactly we need and what are we going to do in the schools (see appendix B).

The teachers who helped in the data collection and supervised the whole training process during the sessions never used the system before, so it was necessary to train them, which required more time and effort and provision of instant help during the data collection.

At the end of the test completions and each of the training sessions, participants received immediate feedback about their achievement. The administration of the tests varied in time regarding the number of tasks included.

For study 1, mouse skills and inductive reasoning tests were administered to the 2nd and 3rd grade pupils during the winter (December 2016 and January 2017). When the results of the tests were good, a decision was made to skip the mouse skills test for two reasons, first, to do the inductive reasoning test on a bigger sample size and to include 4th graders. We also presented the instructions in a written form instead of voice recording and that was the case for the study 2 which we conducted in August/September 2017. Study 3 and 4 were carried out at the beginning of the first semester of the academic year 2018/2019. Well-informed teachers about the eDia system administered the tests and the training. The eDia system provides immediate feedback to the pupils for both cases, the test and the training. The analysis of the results started immediately after the tests were completed.

The data were analysed by SPSS, using the methods of classical test theory, by Mplus using the tools of structural equation modelling and, finally, ConQuest, using the functions of item response theory (IRT). SPSS was used for running the reliability analyses, correlations and finding the differences in performance among pupils in different grades and genders. Independent sample t-tests and analysis of variance (ANOVA) were applied to examine the differences between pupils' achievement according to their grades, gender, school, parents' background information like occupation and level of education, including socio economic status. Cohen's (1988) convention was used for describing the magnitude of effect size (d-index).

IRT was used for scaling the data and visualising pupils' ability level and the items difficulty level on the same scale (item/person maps). "The main idea of item response theory (IRT) is to use a mathematical model for predicting the probability of success of a person on an item, depending on the person's ability and the item difficulty" (Adams &

Wu, 2002, p. 28). IRT models are not deterministic as functions of classical test theory, they express the probability of a correct response to a test item as a function of abilities given one or more parameters of the item.

For building the best fitting measurement model, testing dimensionality of inductive reasoning and running invariance analyses, the tools of structural equation modelling (SEM, e.g. second-order multiple group latent curve modelling) were used (Alessandri et al., 2017). CFI (Comparative Fit Index), TLI (Tucker–Lewis Index) and RMSEA (Root Mean Square Error of Approximation) indices were calculated by MPlus to indicate the model fit.

6 THE EMPIRICAL STUDIES

6.1 Study 1. The feasibility of computer-based testing in Palestine among lower primary school pupils: Assessing mouse skills and inductive reasoning

6.1.1 Introduction

The first study introduces and explores the potential of using computer-based testing at Palestinian schools. It investigates the developmental level of mouse skills and tests the applicability of an online test measuring pupils' inductive reasoning. The study was carried out in the governorate of Bethlehem in Palestine. The government of Palestine gave a big attention to the modern educational technologies in the Palestinian schools by integrating ICT in education. It worked and is still working on implementing several initiatives to develop the educational system towards a more technological education system by investing in the infrastructure of the education system as a whole, schools, teachers and pupils to reach the proposed goals (see Shraim, 2018; PMEHE, 2018; Alhadath, 2016; Shihab, 2014).

We need to explore the applicability and the visibility of computer-based assessment in the Palestinian context in order to meet the main objective of the study. This paper explores the potential of using computer-based testing with young children at the early age of schooling at Palestinian schools. Especially, it investigates the developmental level of mouse skills among grade 2 and 3 pupils and tests the applicability of an online test measuring pupils' inductive reasoning using the ICT facilities of the participating schools. We want to know the factors that influence the performance of the pupils by focusing on the main ones like background factors including the socio-economic status. There are many international assessments i.e. PISA and institutions i.e. OECD which have an impact on the development of CBA. Palestine does not take part in these international programs and studies, investigations in this topic are not visible yet.

6.1.2 Detailed methods of the first study

6.1.2.1 Participants of the first study

The sample members (N = 57) were selected from among second (N = 28) and third grade (N = 29) pupils (age: *Mean* = 7.5; *SD* = .504; 32 male and 25 female) studying in three different elementary schools in Palestine. Pupils at this age are taught and evaluated by class teachers. Pupils have been divided into three levels (advanced 31.6%, intermediate 35.1% and low 33.3%) regarding their achievement at their schools.

6.1.2.2 Instruments administered in Study 1

The study is based on two computer-based tests measuring pupils' mouse (28 items) skills and inductive reasoning (36 items) skills prepared for young pupils, and a questionnaire developed to collect information about pupils' social-economic background and computer use at home.

The mouse skill is a prerequisite to online testing, without a given level of it, computer-based testing is not feasible and valid. Both tests consisted of figural items and in both cases, instructions were given online by a pre-recorded voice.

Pupils' basic computer skills, such as keyboarding and mouse skills were measured in the first test. The test investigated the nature and developmental level of mouse skills by testing two processes: clicking and drag-and-drop. Pupils had to use headsets to listen to instructions then they had to indicate their answer by using the mouse: clicking on the correct picture or pictures or drag-and-dropping the correct picture or pictures (see Figure 6.1). The size and amount of the objects they had to click on or drag-and-drop was changing in the test. In the first tasks they had to click on and drag 1-3 big elements to big places, while at the end of the test they had to cope with clicking on and dragging more than three small elements. Pupils received task-level feedback. If they failed on one of the tasks, they got the task back and had immediately a second chance to do it again. Pupils' responses were scored as correct ("1") if they managed to solve the tasks in the first or second attempt, otherwise, the response was scored as incorrect ("0").

Figure 6.1. A sample item from the mouse skills test (instructions: pupils need to drag the fish into the pool)

The inductive reasoning test comprised two subtests: figural series and figural analogies (see Figure 6.2). In case of figural series, pupils were required to drag the right answer into the yellow box from the shapes below the line that matches the sequence of shapes above the line. In case of figural analogy, pupils were required to find out the missing shape which fits into the yellow box regarding the result of changes in the other two boxes.

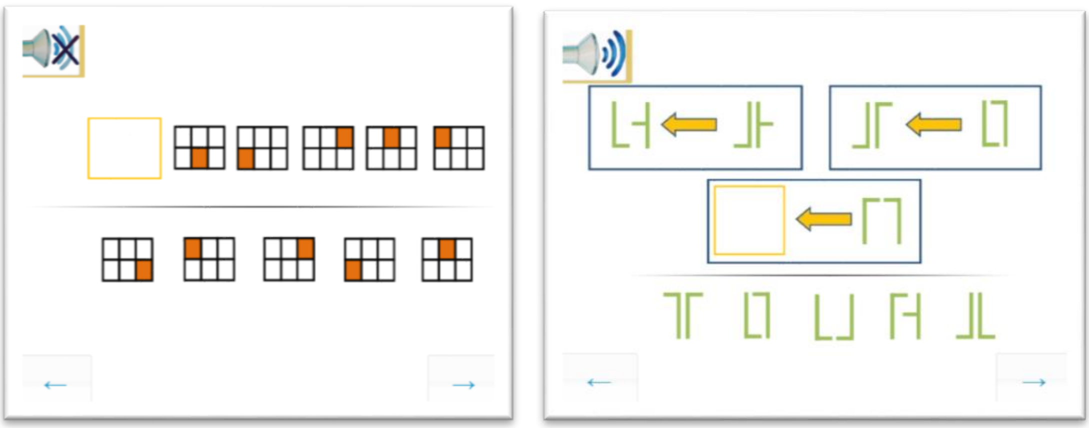


Figure 6.2. Series and analogy items from the inductive reasoning test

The items have been adapted from Hungarian to be suitable to the Arabic style. They have been mirrored by changing the direction of the items (left to right into right to left) to match the Arabic format (see Figure 6.3). The instructions have been translated into Arabic language, recorded and attached to the system; thus they were administered online using headsets. Automatic scoring was used, and instant feedback was provided at the end of the tests. In this case pupils had only one single attempt per tasks, thus pupils' responses were scored as correct ("1") if they managed to give the completely right answer for the first time, otherwise, the response was scored as incorrect ("0").

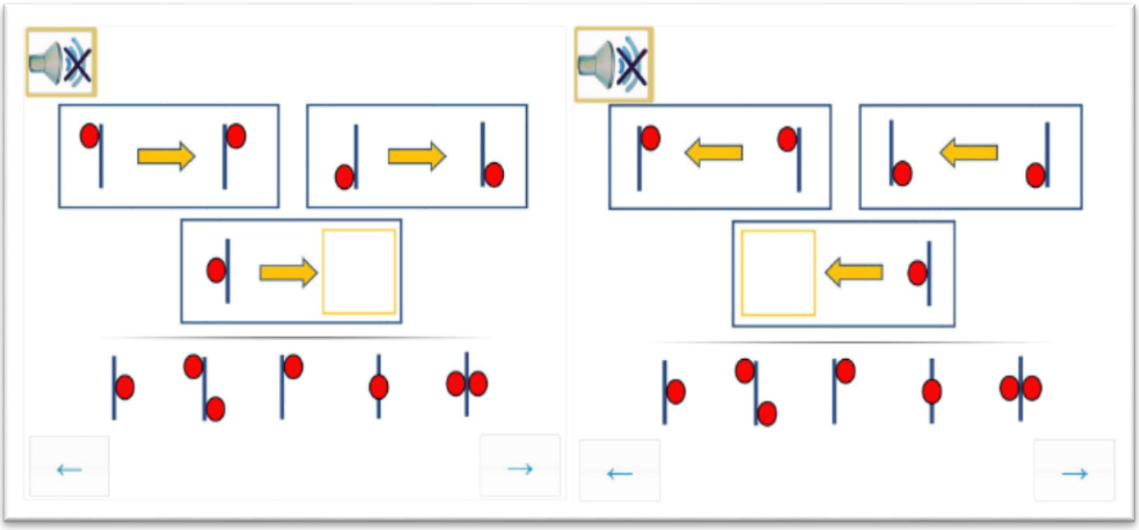


Figure 6.3. Sample item from the inductive reasoning test (left: original Hungarian version, right: translated Arabic version)

6.1.2.3 Detailed procedures of the first study

The online data collection was carried out via the eDia platform using the schools' infrastructure. Test sessions were supervised by teachers who had been thoroughly trained in test administration. At the beginning of the tests, pupils were provided with instructions, through which they could learn how to use the program: (1) at the top of the screen, a yellow bar indicated how far along they were in the test; (2) they had to click on the speaker to be able to listen again to the task instructions; (3) to move on to the next task, they had to click on the "next" button; (4) in the inductive reasoning test pupils received extra warm-up tasks to enhance keyboarding and mouse skills to improve the feasibility of the assessment and validity of the results; and, finally, (5) after completing the last task, they received immediate visual feedback with 1 to 10 balloons, where the number of balloons was proportionate to their achievement. Both tests lasted 45 minutes, one school lesson. Functions of classical and item response theory (IRT) was used by the analyses.

By IRT the individual's response to a specific test item is determined by an unobserved mental attribute of the individual. Both the test items and the individual responding to them are arrayed on a logit scale from lowest to highest. In terms of binary scored test items all IRT models express the probability of a correct response to a test item as a function of abilities given one or more parameters of the item. The Rasch model routinely sets the probability of success at 50% for any person on an item located at the same level on item-person logit scale. The probability of success increases to 75% for an item that is 1 logit easier or decreases to 25% for an item that is 1 logit more difficult.

6.1.2 Results

6.1.2.1 Descriptive statistics and reliability analyses

The reliability of the tests and subtests were examined by computing Cronbach's alpha for each test and subtest. The internal consistencies of the tests were good, with Cronbach's alpha being .75 for the mouse skills test and .92 for the inductive reasoning test. The subtest level analyses regarding IR test revealed that the subtest level results were also generalizable (Cronbach's alpha = .849, .853; EAP/PV Reliability = .865, .865), however, they show an increased probability of measurement errors. The mouse skills test was generally easy (M = 90.53%, SD = 9.67%) for pupils, especially for grade 3 pupils (M = 97.0%, SD = 2.87), whose achievement was significantly higher than second graders' (M = 83.8%, SD = 9.64; t = -7.07, p < .001) mouse skills. The inductive reasoning test was moderately difficult for the pupils at this age (M = 43.46, SD = 23.7). (RQ1)

The learning effect regarding the basic computer skills was supported by the result. Pupils achievement proved to be significantly lower (M = 84.46, SD = 10.89) and the differences between pupils significantly higher without task-level feedback and second trial (t = -7.65, p < .001). Independent from the required procedure tasks requiring the same or similar operations than items somewhere previously in the test were significantly easier than items consisting of the same operation for the first time in the test. These results indicate that pupils mouse skills are generally high even at early years of schooling, and within a short period of time can be effectively enhanced increasing the validity of other achievement tests. Thus taking pupils' computer skills into account, CBA can be used for measuring pupils' performance at this early school age. Pupils' mouse skills are adequate to answer computer-based tests requiring clicking on and drag-and-dropping of big and smaller object or objects. Therefore, the relative lower achievement (Table 6.1) on the inductive reasoning test is not caused by pupils' mouse skills, but their level of inductive reasoning skills. There was no significant correlation between pupils' mouse skills and achievement on the IR test ($r = .215$, $p > .05$). (RQ2)

The test comprised of two subtests measuring inductive reasoning by figural series and figural analogy. Pupils' achievement proved to be the same in both subtest, there were no significant differences ($t = -1.578$, $p > .05$) between the subtest level average achievement. Pupils, whose achievement was high on figural series, tended to achieve high on the figural analogy tasks too ($r = 0.89$, $p < .001$) and on the other way round, pupils, whose achievement was lower on the first subtest, achieved lower on the second one too.

Table 6.1. Descriptive statistics and reliability indices regarding the inductive reasoning test in the first study

Test	Number of items	Mean %	*SD* %	Cronbach's alpha
Inductive reasoning	36	44.15	24.28	.915
Series	18	43.76	25.86	.849
Analogy	18	40.35	22.65	.853

The applicability of the IR test and its subtests were supported by the two-dimensional IRT analyses too. Figure 6.4 shows the match between the item difficulty distribution and the distribution of pupils' Rasch-scaled achievement estimates for inductive reasoning. The two-left panel, headed Fig. Analogies and Fig. Series, show the distribution of pupils' achievement on the two areas of IR, respectively. Pupils at the top end of these distributions had higher achievement estimates than pupils at the lower end of the distributions. The right panel, headed items, shows the distribution of the estimated item difficulties for each of the items on each of the areas. The units of the vertical scale are not scaled, because they are in these Figures arbitrary. In Figure 5.4 every 'x' represents 0.5 pupils.

```
 Fig. analogy  Fig. Series  +item
4             |         |
              |         |
              |         |
              |         |
              |         |
 3           X|         |
              |         |18
              |         |
             X|         |
            XX|        X|
            XX|        X|
 2          XX|       XX|17
           XXX|      XXX|15
            XX|      XXX|
            XX|    XXXXX|
          XXXX|      XXX|14 34
 1      XXXXXX|    XXXXX|31
           XXX|      XXX|13
         XXXXX|     XXXX|10 23 33 35
          XXXX|     XXXX|12
          XXXX|     XXXX|28 36
 0       XXXXX|    XXXXX|16 25 29
          XXXX|     XXXX|8 22 24 26 30
           XXX|   XXXXXX|6 20 32
           XXXXXXXXXXXXX|3 4 27
          XXXX|    XXXXX|1
-1     XXXXXXXXXXXXXXXXX|11
     XXXXXXXXX|  XXXXXXX|21
        XXXXXX|    XXXXX|19
        XXXXXX|      XXX|5 7 9
         XXXXX|     XXXX|2
-2         XXX|       XX|
         XXXXX|       XX|
           XXX|      XXX|
              |       XX|
             X|        X|
              |         |
-3           X|        X|
             X|         |
             X|         |
              |         |
             X|         |
-4            |         |
              |         |
```

Figure 6.4. The two-dimensional item/person map of the inductive reasoning test (every x represents 0.5 pupils)

The items were generally well-matched to the sample (‘x’ and number are almost parallel) but there were items which were relatively hard for the pupils (e.g. item 18) and there were more easy items missing from the IR test. Pupils behaved similar in both dimensions.

Based on the results obtained from the descriptive statistics, reliability analyses and IRT, we can conclude that computer-based assessment is feasible and valid in the Palestinian educational context even at the early age of schooling. The reliability of assessments proved to be good and children generally do not have difficulties handling the computerized tests. The difficulty of the IR test fit the ability level of the sample well, on test and subtest level both, and pupils' achievement on it was independent from their level of mouse skills. (RQ3)

6.1.2.2 Factors influencing the applicability of computer-based assessment of pupils

The frequency of computer use did not influence test achievements (r = -.152, p > .05). The mothers' occupation correlated with their children's IR (r = .285, p < .05), but the mothers' educational background (r = .189, p > .05), or their monthly wage (r = -.254, p > .05) did not influence pupils' achievement. The same results were detectable regarding subtest level achievements. (RQ4)

There were no gender differences detectable in the mouse skill test (M_male = 91.51%, SD = 8.25; M_female = 89.28%, SD = 11.29; t = .83, p > .05). Thus, eventual gender differences on the IR test could not have been caused by the differences in the mouse skills.

In the inductive reasoning test girls achieved significantly higher (M_girls = 50.88, M_boys = 37.67; t = -2.15, p < .05) than boys. This difference was caused by the differences detectable in Grade 2, because there was no significant gender difference detectable in Grade 3 (Table 6.2). (RQ5) (RQ6)

Table 6.2. Grade and gender-level differences in pupils' IR skills

Grade	Gender	Mean	*SD*	*t*	*p*
2	Female	50.23	24.37	-2.15	p < .05
	Male	32.12	20.08		
3	Female	51.49	24.96	-9.20	p > .05
	Male	43.22	23.32		

6.1.2.3 Relationship between pupils' school achievement and the developmental level of IR skills

Pupils with higher achievement at school achieved significantly higher ($M\_school\_advanced = 60.49$, $M\_school\_average = 44.3$; $M\_school\_low = 26.46$, $F = 13.82$, $p < .001$; $r = .58$, $p < .001$) on the IR test. Pupils with higher school achievement proved to be more developed in inductive reasoning skills, thus the development and evaluation of pupils' IR skills are hidden embedded in the Palestinian school curriculum and evaluation process. (RQ7)

6.1.3 Discussion

In this study we raised and identified some important issues concerning the feasibility and the applicability of computer-based assessment and enhancement among young learners in the Palestinian schools. Administering computer-based tests to young children at the first stage of formal schooling may raise several challenges and questions concerning the validity of results.

We confirmed research results from the literature (Molnár & Pásztor, 2015a) that computer-based testing and training can be used with early age pupils, even without modern touch screen computers, using the infrastructure (e.g. desktop computers) provided at schools. The study has shown that both the average hours of using computers at home or outside, and gender did not influence pupils' basic ICT skills and test achievement, thus pupils even at the beginning of schooling are prepared with all those mouse skills (clicking and drag-and-dropping) which are needed to answer the questions appearing in a computerised test.

Several studies have been conducted in different countries concerning whether the stimulation of thinking skills is pursued and evaluated explicitly in schools. In most of the countries, education focuses on reading, writing, and math (Molnár, 2011; de Koning, Hamers, Sijtsma, & Vermeer, 2002) and it is assumed that reasoning skills like inductive reasoning develop spontaneously as a "by-product" of teaching (de Konig, 2000). This view is supported by studies that reported strong correlations between reasoning skills and successful learning of several school subjects, for example, second languages (Csapó & Nikolov, 2009). The findings of the present study suggest that the Palestinian school system supports the explicit development and evaluation of inductive reasoning. Pupils with higher school achievement proved to be more developed, pupils with lower school achievement proved to be less developed in inductive reasoning skills. Thus, the development and evaluation of pupils' inductive reasoning skills must be embedded in the Palestinian school curriculum and evaluation process.

Studies claimed that the background factors like parents' background (mainly the mothers') and socioeconomic backgrounds can influence the achievement of the pupils (see Csapó, 2010; Nikolov & Csapó, 2018; Pásztor, 2016). In our study, we so far go along with the literature regarding the mother's occupational background but not the educational background or the economic status of the families of the pupils. However, it is important to keep in mind that this is a small sample size study, with such a study, it is not possible to fully confirm the achieved results and it is not even generalizable. Therefore, we can claim here that the sample size matters a lot in the accuracy of the results and a bigger one could make a difference.

The gender of the participating pupils is another important factor (see Weaver & Raptis, 2001; Clariana & Wallace, 2002) used for the purpose of comparison in this study. There was diversity in the literature regarding the effect of gender on the achievement of the pupil (Molnár, 2011; Pásztor, 2016; Molnár & Csapó, 2019b; Lőrincz & Molnár, 2012). In this regard, can someone say it is something related to the study environment meaning the school itself, or it might be related to a specific region? Therefore, the best answer for this question was to do this comparison in the Palestinian environment and see its level of effect. In our study, things were close to what we found in the literature, the controversy between the two grades in the inductive reasoning test is also exciting : there were gender differences observed in the second grade but not in the third one. It cannot be assumed that this is something generalisable with the small sample size, but it keeps our study on the same track of controversy as it is in the literature.

Study 1 was one of the first attempts to carry out online assessment in Palestine at elementary school level. Computer-based assessment can be described as a user-friendly instrument for teachers and pupils for monitoring the development of pupils' thinking skills as the results indicated. However, my experience with Palestinian schools suggests that public schools have so far good computer laboratories, but it is not enough to cover all school classes and the number of computers needs to be increased. The internet speed is also another issue where the speed should also be increased to satisfy the needs of the teachers and the computer labs.

The psychometric analysis parameters of the mouse skills test and the inductive reasoning test were in general good for the pupils in Palestine at that age. However, in order to achieve more reliable assessments in this age group we still need to revise the process again with a bigger sample size, especially for the case of the inductive reasoning skills test. Therefore, we cannot claim that the results we have achieved are generalisable, but it can be said that computer-based assessment is applicable in the Palestinian schools and it is worth doing further assessments in the field of thinking skills among school children in Palestine.

Regarding the format of the test we did not detect any difficulties observed by the pupils when doing the test. Therefore, we can assume that the test meets the standards of the format of the Arabic styles of items and tests. Therefore, we will rely on the same format for the future items that need to be added to the test, since the main source of items we usually use is from the Hungarian item bank. In the following section (Study 2), we have implemented the IR test again on a bigger sample size and we also included the fourth graders in the test.

6.2 Study 2. Introducing computer-based assessment among 2nd to 4th grade pupils in Palestine

6.2.1 Introduction

Study 2 presents and tests the applicability of computer-based testing in Palestine by assessing second-, third- and fourth-graders' inductive reasoning skills. It aims to discover background factors like gender, grade and mothers' level of education influencing the applicability of CBA with Palestinian pupils, and to test gender differences regarding inductive reasoning. Computer-based assessment (CBA) is spreading worldwide; education systems are in favour of applying it (Thurlow et al., 2010). Its applicability still raises several questions if it is to assess pupils at the beginning of schooling, but helps pupils tackling different cognitive tasks (Csapó et al., 2014). Therefore, we decided to focus on pupils in their early school age and see if there is any developmental level in the pupils' thinking skills enhanced by their schools. For this purpose, we decided to do the same test as in Study 1 here, and add the fourth-graders to the second and the third ones.

6.2.2 Methods of Study 2

6.2.2.1 Participants

The sample of the study was selected from the second- (N = 61), third- (N = 78) and fourth- (N = 54) graders. In total 193 pupils from four different primary schools in Palestine were included, containing 92 boys and 101 girls. Pupils' age ranged from 7 to 9 years old. Pupils at this age are taught and evaluated by class teachers. Pupils have been classified regarding their achievement at their schools into three levels: advanced, intermediate, and low.

6.2.2.2 Instruments

A computer-based inductive reasoning test was used in this study. It consists of multiple-choice figural items in two main types: series and analogy (in total 33 items). The items were adapted from the original Hungarian item bank (Molnár et al., 2013; Pásztor et al., 2017) and a questionnaire was developed to collect information about pupils' socio-economic background and their average computer use.

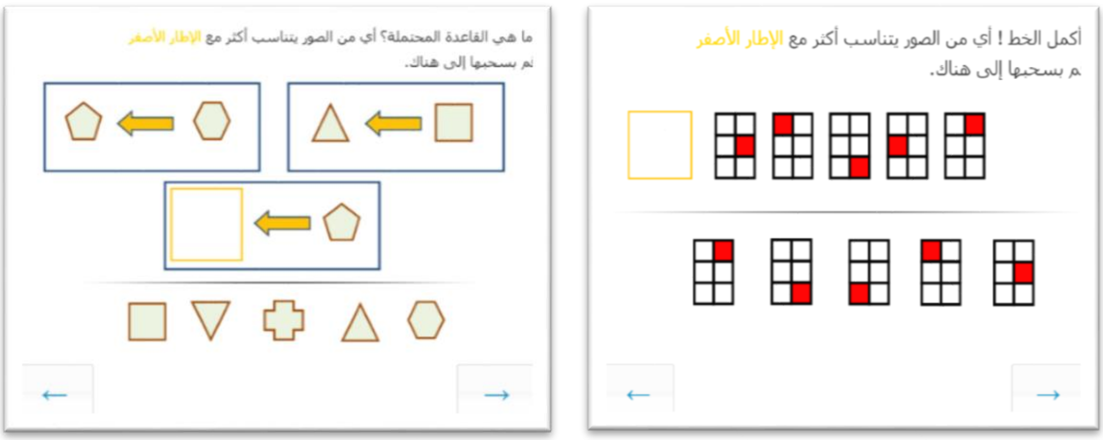


Figure 6.5. Sample items of the figural series and figural analogy subtest of the inductive reasoning test

The IR test comprised of two subtests. The first one contained figural series items, where pupils were asked to drag the right answer into the yellow box from the shapes below the line that matched the sequence of shapes above the line. The second subtest consisted of figural analogies, where pupils had to figure out the missing shape which fitted into the yellow box regarding the result of changes in the other two boxes (see Figure 6.5).

We have adapted the original Hungarian items to Arabic format to minimize any obstacles that might be related to the format of the test. For this purpose, the original Hungarian items were mirrored by changing the direction of all items to suit the Arabic style. The texts of the instructions were translated to simplified Arabic and it was added to the system (see Figure 6.6). For this test, the pupils received the instructions about the tasks in form of written texts. The eDia system supports automatic scoring and it also provides instant feedback at the end of the test. Pupils had only one single attempt per tasks, that means if the pupil managed to answer correctly, the system scored as correct ("1"), if not, then the system scored it as incorrect ("0"). After that, the pupil was directed to the next task immediately.

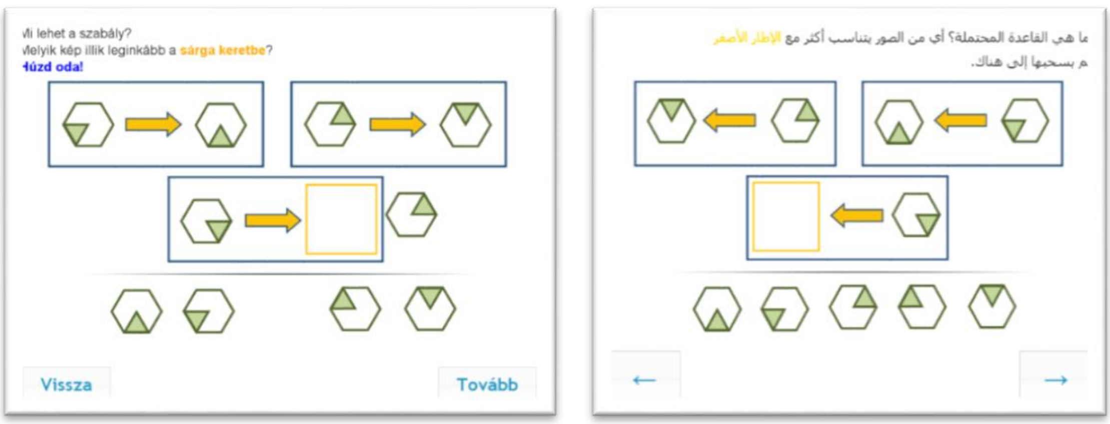


Figure 6.6. The same item from the original Hungarian and the Arabic version of the inductive reasoning test

5.2.2.3 Procedures and data analysis of Study 2

The online data collection was carried out via the eDia platform (Molnár & Pásztor, 2015a) using the schools' available infrastructure of computer labs. Test completion lasted approximately forty-five minutes. Teachers who received full information about the test and its administration supervised the sessions completely. The test started with providing the pupils with general instructions about the progress of the test and the way it works. The first page of the program was an entrance page with password login, then the second page included instructions about the test in general, after that pupils started the test with a warm up task (with no credits for it), each task included instructions at the top of it, in addition, there was an orange line indicating how far the pupil was in the test. When the pupil reached the end of the test, a percentage evaluation appeared immediately after the last task of the test giving the examinee instant feedback. Pupils' took approximately 45 minutes to complete the whole test. For the aim of the analysis, two main models were used to analyse the results which were functions of classical (Cronbach-alpha, bivariate correlation, t-test, mean, standard deviation) and item response theory (IRT).

Rasch analyses were used to test the difficulty of the tasks and find out about their appropriateness level for the pupils. The diagram shows the ability measure of the pupils on the left side and the items' difficulty distributions on the right side. The items that were more difficult for the pupils are located higher on the scale then followed by the less ones below it. The higher the ability level of a pupil is, the higher it is located on the scale, then followed by a lower ability level pupil and so on. The pupils are usually represented by an "x" in the scale and each "x" represents a specific number of pupils, depending on the participants' numbers and the number of the items.

6.2.3 Results of Study 2

6.2.3.1 Descriptive statistics and reliability analyses

The reliability of the adapted test was measured using the Cronbach's alpha coefficient for the test and the subtests as well. The internal consistencies of the test were good, with a high Cronbach's alpha for the IR test (α = .90). The pupils were able to finish the test on time and answer all questions (tasks) that required the participants to use the basic mouse skills (drag-and-drop and clicking). Therefore, when the basic mouse skills were applied on IR tasks in a computer-based testing, pupils were able to do these tasks effectively and efficiently at the age between 7 and 9 years old. (RQ8)

The difficulty level of the inductive reasoning test was moderately average for the pupils at this age based on the pupils' mean and row score distribution (*Max* = 94.44, *Min* = 51.84, *M* = 51.84, *SD* = 21.88). The subtest analysis of the test shows that both figural series and figural analogy results are also generalizable and the identification of similarities or differences, and dissimilarities in a series or analogies were easy for the pupils in case of figural series and figural analogy (see Table 6.3).

Table 6.3. Sub-test level descriptive statistics of the IR test in Study 2

Test	Number of items	Mean %	*SD* %	Cronbach's alpha
Series	17	56.24	23.51	.83
Analogy	16	47.43	23.64	.85

The item/person map (Figure 6.7) shows the item difficulty distribution and the distribution of pupils' Rasch-scaled achievement estimates for the IR test. The "x" to the left side represents the pupils' distribution, at the top are located the pupils with higher achievement estimates than pupils at the lower end of the distributions. The right side shows the distribution of the items regarding its difficulty where the more difficult items are located on the top of the panel. The figure shows that most items are in the middle and pupils were mostly located around these items which means that the items are generally well-matched to the sample distribution.

It can be concluded, based on the results obtained from the descriptive statistics, reliability analyses and IRT, that computer-based assessment is feasible and valid in Palestinian schools among early age pupils (7-9 years old). Pupils did not have any difficulties dealing with computerised tests and the reliability of the assessment proved to be good. The IR test difficulty is standard for the ability level of the pupils in the test and subtest level. (RQ9)

```
   4                                          |                                        |
                                              |                                        |
                                              |                                        |
                                             X|                                        |
                                              |                                        |
                                              |                                        |
   3                                          |                                        |
                                             X|                                        |
                                           XXX|                                        |
                                          XXXX|                                        |
                                         XXXXX|                                        |
                                   XXXXXXXXXXX|                                        |
                                      XXXXXXXX|                                        |
   2                               XXXXXXXXXXX|14                                      |
                             XXXXXXXXXXXXXXXXX|15                                      |
                       XXXXXXXXXXXXXXXXXXXXXXX|16                                      |
                         XXXXXXXXXXXXXXXXXXXXX|                                        |
                          XXXXXXXXXXXXXXXXXXXX|                                        |
                       XXXXXXXXXXXXXXXXXXXXXXX|10 17 34                                |
                      XXXXXXXXXXXXXXXXXXXXXXXX|                                        |
   1         XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                        |
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                        |
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|8 11 18                                 |
                   XXXXXXXXXXXXXXXXXXXXXXXXXXX|12 27                                   |
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                        |
          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|25 28 31 35                             |
   0                 XXXXXXXXXXXXXXXXXXXXXXXXX|13 23 33                                |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|9 24 26 30                              |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|36                                      |
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|32                                      |
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|1 3 4 7                                 |
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|5 29                                    |
          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|20 22                                   |
  -1                      XXXXXXXXXXXXXXXXXXXX|6                                       |
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                        |
              XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|21                                      |
                              XXXXXXXXXXXXXXXX|                                        |
                                XXXXXXXXXXXXXX|2                                       |
                                         XXXXX|                                        |
                                     XXXXXXXXX|                                        |
  -2                                       XXX|                                        |
                                           XXX|                                        |
                                             X|19                                      |
                                              |                                        |
                                             X|                                        |
                                            XX|                                        |
  -3                                        XX|                                        |
                                             X|                                        |
                                              |                                        |
```

Figure 6.7. The item/person map of the second study (each x represents 0.3 pupils)

6.2.3.2 Grade-level differences

Pupils in grade 3 achieved significantly higher than their mates in grade 2 (*M_grade2* = 41.48, *SD* = 19.79; *M_grade3* = 57.65, *SD* = 26.04; $t = 3.70$, $p < .01$). No significant differences were detected between grade 3 and 4 (*M_grade3* = 57.65, *SD* =

26.04; *M_grade4* = 63.07, *SD* = 18.93; $t = 1.38$, $p > .05$), and strong significant difference occurred between grade 2 and 4 pupils ($t = -5.49$, $p < .01$).

Figure 6.8 shows the distribution curves of grade 2-4 pupils' achievement in the IR test. Thus it can be noticed, that at this period of age, pupils' inductive reasoning skills are being enhanced by growing up in age at their schools without being subjected to any training programs from outside to affect that development. (RQ10)

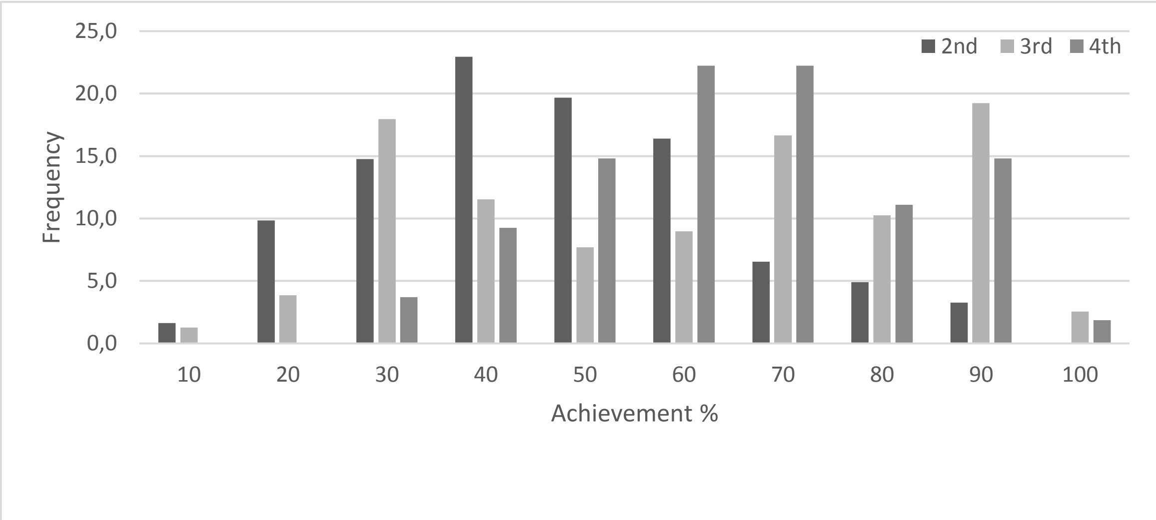


Figure 6.8. Grad-level distribution curves of IR

6.2.3.3 Gender differences

No significant gender-level differences were detected in both samples (*M_female* = 54.70, *SD* = 20.64; *M_male* = 48.70, *SD* = 22.88; $t = -1.92$, $p > .05$) and grade levels (see Table 6.4). Girls and boys achieved equal, at the same level in the same grade. Figure 6.9 shows the distribution curves of both genders in the IR test. (RQ11)

Table 6.4. Grade- and gender-level differences in pupils' IR skills

Grade	Gender	Mean	*SD*	*t*	*p*
2	Female	44.87	17.52	-1.62	$p > .05$
	Male	37.45	18.52		
3	Female	56.38	23.11	-.654	$p > .05$
	Male	52.77	25.63		
4	Female	63.21	16.25	-1.61	$p > .05$
	Male	55.55	18.54		

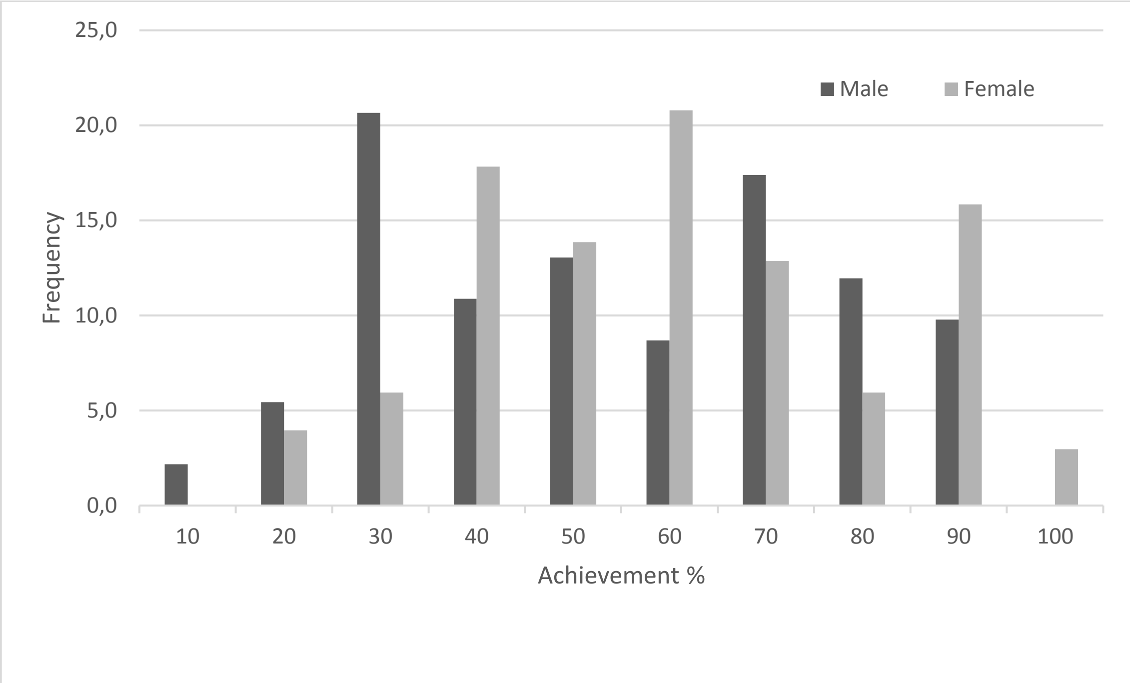


Figure 6.9. Gender-level distribution curves of the IR test

The background variables that could play a main role in influencing pupils' inductive reasoning skills in the Palestinian school children were the level of the mother' education and her occupation. Therefore, we mainly focus on these two factors regarding their importance. The results confirmed and proved that pupils' achievement in the IR test is strongly influenced by the educational background of the mother (r = .694, p < .01); the mother's occupational background also influenced strongly the achievement of the pupils (r = .539, p < .01). Pupils have higher level of IR skills when their mothers are more educated and vice versa, when mothers have less education, pupils' level of IR is lower. Pupils' school achievement had also a strong relationship with their test-level achievement (r = .869, p < .01), it means that pupils who achieved higher in their school were more developed in inductive reasoning skills, thus the development and evaluation of pupils' IR skills are available in the Palestinian school curriculum. The frequency of computer use did not influence test achievements (r = -.024, p > .05). (RQ12)

Low school achievers achieved significantly lower and high school achievers achieved significantly higher on the test than average achievers in all grades (*M_school_advanced* = 78.28, *M_school_average* = 55.67; *M_school_low* = 30.01, F = 95.34, 166.17, 129.01; in all cases p < .001). The bivariate correlation test indicated a strong relationship between school achievement and the developmental level of inductive reasoning in all grades (r = .087, .899, .913; p < .001). We can say regarding the results that pupils with higher level reasoning skills were from the top performers based on their school marks, while pupils with lower level thinking skills were from the low performers even at school level. This result confirms that the development and evaluation of thinking skills, especially inductive reasoning skills are present implicitly in the Palestinian school curriculum. (RQ13)

6.2.4 Discussion

In Study 2 we investigated the visibility and the applicability of computer-based assessment in the Palestinian schools, among the second-, third- and fourth-graders. Challenges regarding administering computer-based tests become much bigger when it comes to pupils in their early school ages. Teachers need more training to be able to deal with children at that age, to be able to administer the computerised tests in an efficient way.

We confirmed the results from literature (Molnár & Pásztor, 2015a) that supports the applicability of computer-based testing and training with elementary school pupils using the new technological devices at their schools, mainly desktop computers. This study also confirmed our first small sample size results in Study 1 (Mousa & Molnár, 2019a) that computer-based testing can be used in Palestine and it can work effectively even when it comes to early age pupils. Therefore, the online assessment instrument for the thinking skills proved to be reliable regarding the whole test. We also assumed that the development and the assessment of inductive reasoning is embedded in the Palestinian school curriculum, and that can be seen in the pupils’ performance where the IR skills develop as the pupils grow up in age at their schools.

Pupils’ basic ICT skills and test achievement was not influenced by the average hours of using computers at home or outside and gender, that means pupils know the basic mouse skills like drag-and-dropping and clicking well, which the test is mainly built on to answer its tasks by performing these skills.

Several issues were raised in some studies (Mousa & Molnár, 2019a) regarding the background factors, socioeconomic backgrounds, the gender, and the accuracy of the results with a small sample size. We could not agree more that a bigger sample size provides more accurate and precise results. In case of the mother’s educational and occupational backgrounds, both influenced the achievement of the participating pupils in this study, but that was not the case in study 1 for the educational background. In this regard, study 2 goes along with the literature (see Csapó, 2010; Nikolov & Csapó, 2018; Pásztor, 2016) that these factors play a role in the pupils' achievement. Moreover, it can be added that the bigger the sample is, the better the results are. Gender is also a point of discussion among researchers since it can affect the pupils' achievement in some cases (Molnár, 2011; Pásztor, 2016; Mousa & Molnár, 2019a; Molnár & Csapó, 2019b; Lőrincz & Molnár, 2012). In our case (regarding the Palestinian context) it did not have any influence but its importance in the comparison still existed.

The inductive reasoning test proved to be reliable, hence the decision was made to use it again for future assessments. The test consists of two types of items; figural series and figural analogy. The test needs to be developed to cover more important types of items, like numerical items as well.

6.3 Study 3. Applying computer-based testing in Palestine: Assessing fourth- and fifth-graders' inductive reasoning

6.3.1 Introduction

Technological tools become a vital part of our daily lifestyle. Children are the most affected and attracted part of the society when it comes to technology. It does not affect the way they live and play only but also the way they learn and gain knowledge since technology is present in the field of education in the form of e-learning and edutainment. Implementing technology in education has many advantages over traditional methods. It is considered as a modern and attractive way regarding the given value, as it is in computer games which cannot be found in the traditional tools (Bottino et al., 2007). Such features make scholars suggest to develop computerised testing to have new and modern ways of educational evaluation (Csapó, Lörincz & Molnár, 2012) and improve the assessment (Csapó, Ainley, Bennett, Latour & Law, 2012).

In Study 3, we explored the possibility of using computerized tests with pupils in an early school age by assessing IR skills. An online test was used to collect data by measuring fourth and fifth grade pupils' inductive reasoning (IR) skill. Pupils were tested in their own schools using the school's facilities.

6.3.2 Methods

6.3.2.1 Participants of Study 3

The sample of the study was selected from 4th and 5th graders (aged ± 9-10) in Palestinian primary schools. There were 248 pupil participants in this study, 138 boys and 110 girls. Based on their school achievement, pupils were classified into three groups: low, intermediate and high performers.

6.3.2.2 Instruments

A computer-based inductive reasoning test was used in this study. Different formats with different levels of difficulty of inductive reasoning items were administered via the online assessment platform eDia (Molnár & Csapó, 2019a). The test consisted of 60 items in total divided into four different subtests: figural series (18 items), figural analogies (18 items), number series (16 items) and number analogies (8 items). Pupils completed tasks by moving objects (figures or numbers) on the screen by the drag-and-drop function. Sample items of the inductive reasoning test are presented in Figure 6.10. The pupils were asked to identify and point out relationships, similarities or differences, and dissimilarities in a series or analogies, between groups of figures or numbers.

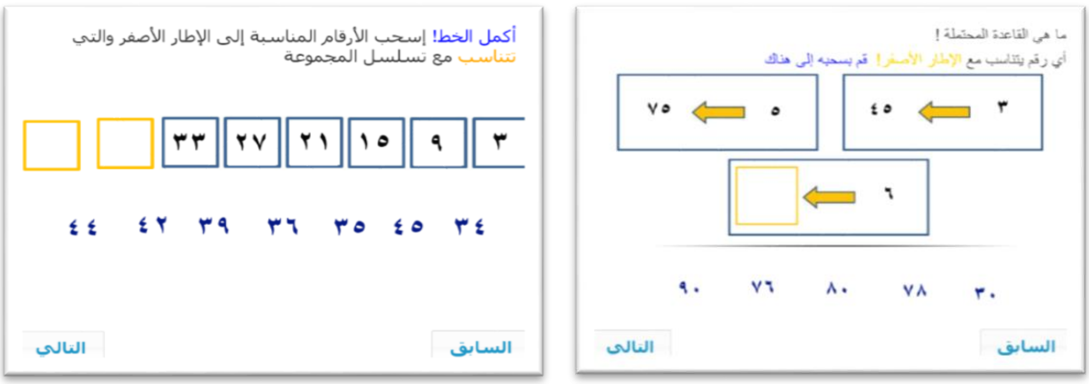


Figure 6.10. Examples of numerical series and numerical analogy items.

In both the figural and number series tasks pupils had to figure out the rule behind the task, then drag the right answer below the line on the screen to complete the sequence. In case of figural and number analogies, pupils also needed to find out the missing shape or number that fitted into the empty box regarding a rule. The rule could be figured out in each of the tasks unequivocally.

The items of the test were selected and adapted from a scaled item bank for inductive reasoning developed in Hungary. Later, all items were adapted to be suitable for the Arabic style as follows: the direction of the items were changed according to the Arabic way by mirroring them (see Figure 6.11), the instructions were translated into the simplified Arabic language and attached to the system in a textual form. The tasks were scored automatically. Pupils' responses were scored as "1" if it was fully correct; otherwise, the response was scored as incorrect ("0"). The pupils received instant feedback at the end of the test.

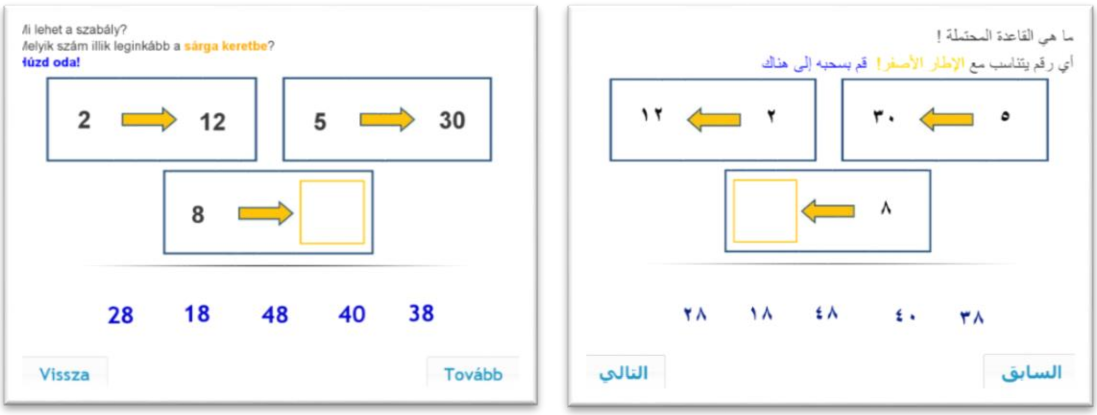


Figure 6.11. Same item from the original Hungarian and the Arabic version of the inductive reasoning test

6.3.2.3 Procedures in Study 3

The test was administered in equipped computer labs using the online eDia platform. At the beginning, participants got instructions about handling the user interface that included also a warm-up task with the aim of improving the feasibility of the assessment and validity of the results. The instructions were given in written format. A yellow progress bar at the top of the screen indicated pupils' actual status in the test, that is, how far they were in the test.

Pupils' inductive reasoning performance was automatically scored, thus, they received immediate performance feedback at the end of the testing session. The test took approximately one hour to complete. Testing sessions were supervised by classroom teachers who have been fully trained in test administration. Both the functions of classical test theory (Cronbach-alpha, bivariate correlation, t-test, mean, standard deviation) and item response theory (IRT) were employed by the analyses.

Rasch analyses were used to test the appropriateness of the tasks regarding their difficulty level. The four-dimensional item/person map shows how the distributions of pupils and items in each of the subtests relate to one another by locating both items and pupils on the same scale. The ability measure of pupils is on the left side of the figures, while the difficulty distributions of the items in each of the dimensions are on the right. More difficult items are positioned higher on the scale than less difficult ones. Pupils with a higher ability level are positioned higher on the same scale then pupils with a lower ability level. Pupils and items are located at the same level if the ability level of the pupil is equal to the difficulty level of the item, that is, if the pupil has a 50% chance of answering the item correctly.

6.3.3 Results

6.3.3.1 Descriptive statistics and reliability analyses

The internal consistencies (Cronbach's alpha) of the inductive reasoning test was good (α = .807), that is, the results of the study are reliable and generalizable. Pupils were able to finish the test on time and answer the questions using drag-and-drop operations with the mouse. Thus, computer-based testing is applicable at the age of 9-10 in the Palestinian school system, when drag-and-drop items are used. (RQ14)

Based on participants' mean and row score distribution, the test proved to be difficult for the pupils at this age (*Min* = 6.67, *Max* = 53.33, *M* = 25.29, *SD* = 10.94). The subtest-level analyses indicated that the identification of similarities or differences, and dissimilarities in a series or analogies were significantly easier if it was about figures and not numbers. The same operations proved to be harder on an average if the content changed to numbers, thus, items having mathematical context and requiring counting proved to be much harder, especially, if they belonged to the number series subtest (see Table 6.5).

Table 6.5. Sub-test level descriptive statistics of the IR test in Study 3

Sub-tests	Items	Min	Max	Mean	*SD*
Figural Analogy	18	.00	83.33	32.63	17.72
Figural Series	18	.00	83.33	41.30	19.01
Number Analogy	8	.00	87.50	16.63	15.15
Number Series	16	.00	18.75	3.35	4.99

According to the four-dimensional item/person map (Figure 6.12) there were big differences in pupils' achievement in the four dimensions of IR. The achievement-based distribution curves were very similar in case of the figural items, which difficulty level fit to the ability level of the pupils. But they strongly differed by items based on number series, which were too hard for the target population, for the 4th and 5th graders. However, the distribution of the items according to their difficulty was good in each of the dimensions, that is, there were easy, medium, and hard items in each of the dimensions. (RQ15)

```
       Num. An. Num. Ser. Fig An. Fig Ser.    +item
--------------------------------------------------------------------------------
  2             |         |         |         |
                |         |         |        X|
                |         |         |        X|32
                |         |         |         |57
                |         |         |        X|10 12 15 17 23 49 58
  1             |         |         |       XX|39 40 55
                |         |        X|      XXX|5 6 41 59
                |         |       XX|     XXXX|9 20 36 46 47
                |         |      XXX|    XXXXX|38
                |         |      XXX|   XXXXXX|4 52
  0             |         |    XXXXX|  XXXXXXX|2 8 33 34 37
                |         |    XXXXX|  XXXXXXX|3 11 16 27 31 45 48
              XX|         |   XXXXXX|   XXXXXX|13 14 18 26 28 53 56
              XX|         | XXXXXXXX| XXXXXXXX|35 44 60
             XXX|         | XXXXXXXX|  XXXXXXX|24 50
 -1         XXXX|         | XXXXXXXX|   XXXXXX|1 7 29 42
          XXXXXX|         |XXXXXXXXX|   XXXXXX|30
         XXXXXXX|         |   XXXXXX|   XXXXXX|25 43 51 54
      XXXXXXXXXX|         |  XXXXXXX|      XXX|
      XXXXXXXXXX|         |   XXXXXX|      XXX|
 -2   XXXXXXXXXX|         |     XXXX|        X|
       XXXXXXXXX|         |       XX|        X|
          XXXXXX|        X|       XX|        X|22
         XXXXXXX|        X|        X|        X|
           XXXXX|       XX|         |         |
 -3          XXX|       XX|         |         |
              XX|      XXX|         |         |
               X|       XX|         |         |21
                |      XXX|         |         |
                |     XXXX|         |         |
 -4             |    XXXXX|         |         |
                |    XXXXX|         |         |
                |  XXXXXXX|         |         |
                |  XXXXXXX|         |         |
                |  XXXXXXX|         |         |
 -5             |  XXXXXXX|         |         |
                |    XXXXX|         |         |
                |      XXX|         |         |
                |    XXXXX|         |         |
                |    XXXXX|         |         |
 -6             |     XXXX|         |         |
                |       XX|         |         |
                |       XX|         |         |
                |       XX|         |         |
 -7             |         |         |         |
                |         |         |         |
                |        X|         |         |
                |         |         |         |
                |        X|         |         |
 -8             |         |         |         |
```

Figure 6.12. The four-dimensional item/person map (each x represents 3 pupils; blue: num. analogies, green: number series, orange: figural analogies, red: figural series).

6.3.3.2 Grade- and gender-level differences

Pupils in grade 5 achieved significantly higher than their mates in grade 4 (M_grade4 = 23.88, SD = 10.26; M_grade5 = 26.68, SD = 11.44; t = -2.02, p < .05) (see

Figure 6.13). The statistics in the figure below show the distribution curve of the inductive reasoning test where the highest frequency of the achievement can be seen around thirty percent, and the fifth grade pupils' frequency is higher compared to the fourth one's when the achievement goes up. This indicates that the age of 9-10 could be a sensitive period for enhancing pupils' inductive reasoning skills, as even without explicit training significant development occurred. (RQ16)

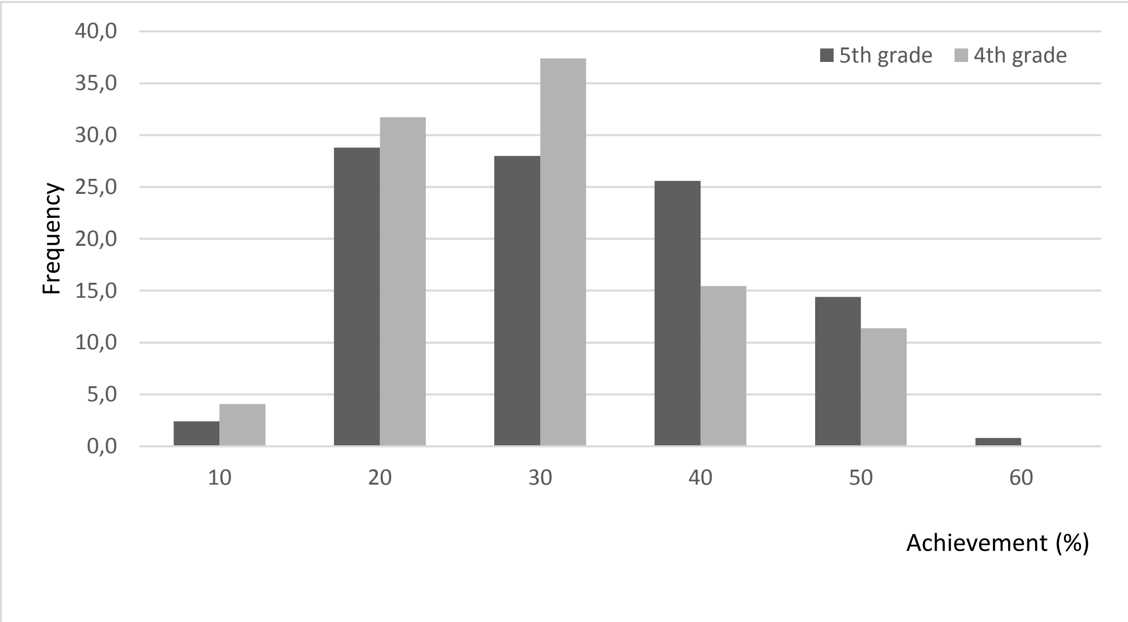


Figure 6.13. Grad-level distribution curves of IR

No significant gender-level differences were detected in both samples (*M_girls* = 26.75, *SD* = 9.81; *M_boys* = 24.13, *SD* = 11.66; *t* = -1.92, *p* > .05) and grade levels (see Table 6.6). Male and female pupils achieved at the same level in the same grade. The statistics in the table below show that the mean and the standard deviation level between boys and girl in the same grades, the fourth and the fifth, are so close. However, when it comes to more detailed level, there can be minor differences, but there are no mean differences according to the distribution curves in Figure 6.14. (RQ17)

Table 6.6. Grade- and gender-level differences in pupils' IR skills

Grade	Gender	Mean %	*SD*	*t*	*p*
4	Female	24.78	9.20	-.85	$p > .05$
	Male	23.18	11.03		
5	Female	28.66	10.08	-1.79	$p > .05$
	Male	25.07	12.27		

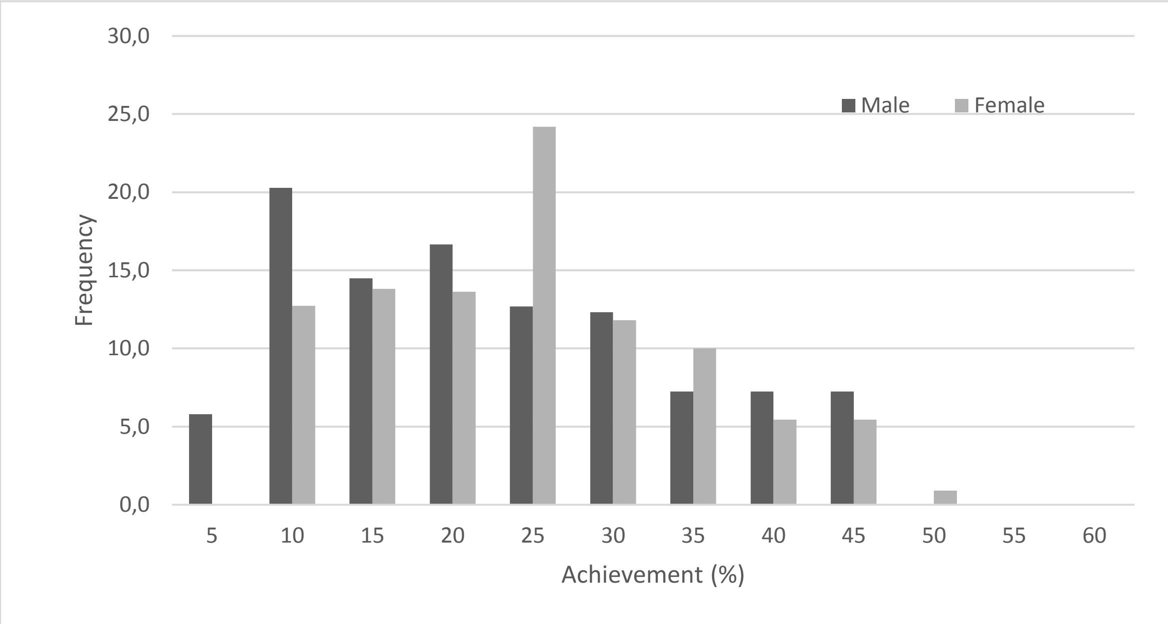


Figure 6.14. Gender-level distribution curves of IR

There are several factors that could influence pupils' level of inductive reasoning skills. In this regard, we focused mainly on the educational and occupational level of the mother, since she plays an immense role in the education of the Palestinian children. The results confirmed and proved that the mother's educational background is a strong influential factor on pupils' achievement in the IR test ($r = .691$, $p < .01$); the mother's occupational background also influenced strongly the achievement of the pupils ($r = .470$, $p < .01$). Pupils with more educated mothers had higher level of inductive reasoning skills than pupils with less educated mothers. Pupils' school achievement had also a strong relationship with their test-level achievement ($r = .897$, $p < .01$), that is, pupils with higher school achievement proved to be more developed in inductive reasoning skills, thus the development and evaluation of pupils' IR skills are hidden embedded in the Palestine school curriculum and evaluation process.

Pupils' level of inductive reasoning had a strong effect on their school achievement (*M_school_advanced* = 40.08, *M_school_average* = 25.48; *M_school_low* = 14.50, $F = 528.16$, $p < .001$; $r = .89$, $p < .001$). Pupils with higher level reasoning skills belonged to the top performers based on their school marks, while pupils with lower level thinking skills belonged to the low performers even at school level (see Figure 6.15). This result also confirms that the development and evaluation of thinking skills, especially inductive reasoning skills, are present implicitly in the Palestinian school curriculum. (RQ18)

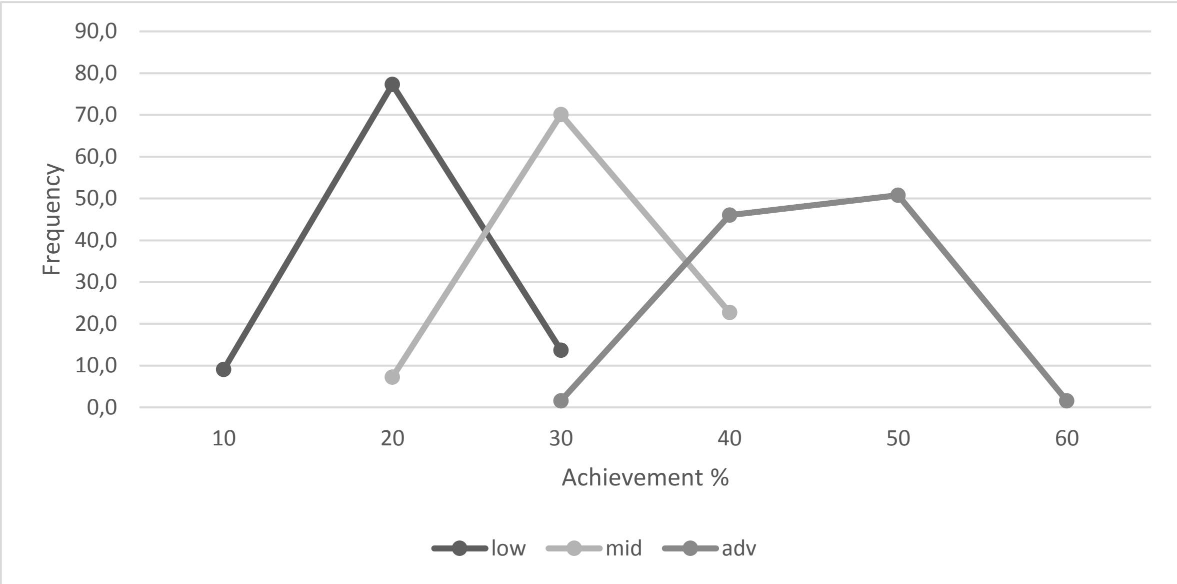


Figure 6.15. Pupils' achievement in the IR test regarding their achievement in their schools in percentage

6.3.4 Discussion

In Study 1–3 (see Mousa & Molnár, 2019a, 2019b) the issue of the applicability of computer-based assessment was raised and confirmed in Palestinian school practice. In Study 3 – next to the applicability of CBA – the main focus was on the development of an IR test, which has good psychometric indexes and can be applied as a pre- and posttest in Study 4, for diagnosing the effect size of the inductive reasoning training.

According to the results we managed to develop an inductive reasoning test, which can be applied in everyday school practice in Palestine, even when it comes to pupils in their early age, using desktop computers provided by their schools, and which has good psychometric indices. The reliability of the test was $\alpha = .807$. The mean and row score distribution of the participants was (*Min* = 6.67, *Max* = 53.33, *M* = 25.29, *SD* = 10.94).

The grade achievement was (*M_grade4* = 23.88, *SD* = 10.26; *M_grade5* = 26.68, *SD* = 11.44; $t = -2.02$, $p < .05$) and the gender achievement was (*M_girls* = 26.75, *SD* = 9.81; *M_boys* = 24.13, *SD* = 11.66; $t = -1.92$, $p > .05$).

We have previously raised a question to discuss whether the stimulation of thinking skills is pursued and evaluated explicitly in schools. We noticed in the analysis of the results of study 1 (Mousa & Molnár, 2019a) that there were development in pupils' performance when third-graders performed better than the second graders. Later in the findings of the 2nd Study, we confirmed what we found in the analysis of the results, that the fourth-graders performed better than the third-graders. In Study 3, we found that the fifth grade pupils performed better than the second-graders, and that proved the Palestinian school system supports the explicit development and evaluation of inductive reasoning. However, the results of the mean performance indicated that there were no significant differences between genders in the IR test. That goes along with the previous studies which indicated the same results (See Weaver & Raptis, 2001; Clariana & Wallace, 2002; Molnár, 2011; Pásztor, 2016; Molnár & Csapó, 2019b; Csapó, Lőrincz, & Molnár, 2012; Mousa & Molnár, 2019a) and it was the same as in study 2.

We confirm the results detected in some studies regarding the relationship between parents' education level and academic achievement (see Asad khan, Iqbal, & Tasneem, 2015). The studies emphasised the significant positive relationship: the higher degree the parents have (mainly the mother), the better the achievement the pupil is (see also Csapó, 2010; Nikolov & Csapó, 2018; Pásztor, 2016). We could not agree more with the literature about that when we noticed the same in our study. One might generalise these results even with some controversial results noticed in other studies, as it was with Kambeyo (2017) when he found no relationship between patents' level of education and pupils' achievement. At the same time, the researcher suggested that economic reasons might cause such a result. That means we cannot consider that as another different result without discriminating any other factor which might play a role in affecting the results.

After the positive results we received for all the previous studies regarding the applicability of computer-based assessment in the Palestinian schools, we have decided to go for applying the online enhancement program of thinking skills on pupils in the Palestinian schools.

6.4 Study 4. Computer-based training in Maths improves inductive reasoning of 9- to 11-year-old children

6.4.1 Introduction

One of the major challenges in classroom teaching comes from the large differences between pupils in terms of abilities. Training programmes are needed which can be used even on the classroom level while handling individual differences, fitting the actual needs determined by the pupils' cognitive level (Pásztor, 2016). Technology-based training programmes can provide feasible solutions to address this challenge. Beyond adaptive fostering, technology has the power to provide higher-level motivation for learners of the 21st century.

We address in this study a technology-based training programme of inductive reasoning for 9–11-year-old pupils and present the immediate results of an evaluation study. The online training consisted of 120 playful problems based on Klauer's "Cognitive training for children" concept and on his theory of inductive reasoning (Klauer, 1989). All the problems were embedded in mathematical content, so the training tasks were applicable during normal school hours as part of the mathematics lesson. To our knowledge, there is no training programme available online in Arabic which is empirically proved and focuses on the development of pupils' inductive reasoning skills in an educational context.

6.4.2 Methods of Study 4

6.4.2.1 Participants

The sample for the study was selected from fourth- and fifth-grade pupils (aged 9–11) in four Palestinian primary schools. We wanted the age group to be as much close as possible to the best age group suggested by researchers (11-13) to develop reasoning skills (see Molnár et al., 2013; Wu & Molnár, 2018a). A total of 236 pupils participated in the study: 118 pupils were assigned to the experimental group and 118 to the control group. The experimental group consisted of 57 fourth-graders and 61 fifth-graders, while the control group consisted of 60 fourth-graders and 58 fifth-graders. In the participating schools, like schools in Palestine generally, each grade is made up of several forms. Every school year, pupils are rearranged between the forms based on their school grades to achieve a balanced distribution of low, intermediate, and high achievers within each form. Forms are taught the same way, using the same teaching methods. As entire forms were selected for the study, members of both the experimental and control groups were taught for the same number of school hours during the period when the training was administered. The retention rate was 0% in both groups. Regards school achievement, 31.4/37.3/31.4% of the pupils in the experimental group and 33.4/44.9/22.0% in the control group were low/intermediate/high achievers, respectively. The mother's educational attainment was approximately the same in the two groups (see Table 6.7).

Table 6.7. The distribution of the sample based on the mothers' educational attainment

	Control group (%)	Experimental group (%)
Undereducated	12.7	12.7
Primary school	24.6	25.4
High school certificate	27.1	22.0
Diploma (A levels)[4]	31.4	28.0
BA/BSc and MA/MSc	4.2	11.9

6.4.2.2 Instruments

Similar to Klauer's original programme, the training consisted of 120 learning tasks in total, with 20 problems for each class of inductive reasoning (generalisation, discrimination, cross-classification, recognising relations, discriminating between relations, and system formation). All the learning tasks could be completed through appropriate inductive reasoning processes, and all were embedded in various mathematical content corresponding to the targeted age group: even and uneven numbers, Roman numerals, relationship between numbers and quantities, fundamental operations of arithmetic, use of relational math symbols, measurements, conversion of units of measurements, series (completing and ordering), data pairs, correlations among triplets, concepts in geometry, geometric transformations, measurement of time, and knowledge of the clock.

The training tasks contained pictures appropriate for young pupils, who indicated their answers with a mouse. The operations were based exclusively on clicking and drag-

[4] In the Palestinian education system, there exist a lower university or college certificate after 2 years of study, which is called diploma. Bachelor requires 4 years of study.

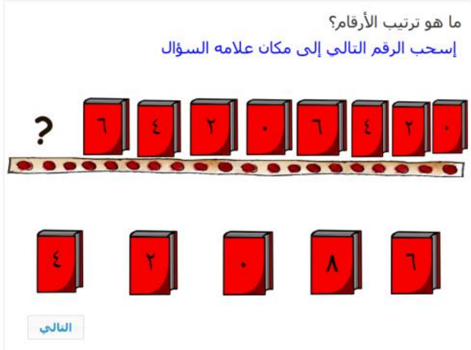
and-drop from left to right. Based on earlier research results (Mousa & Molnár, 2019), the size and number of objects on which they clicked or dragged and dropped did not influence the success and difficulty of the training tasks significantly.

According to Klauer's framework (Klauer & Phye, 2008, p. 88), we have used the following inferential formats according to the six IR operations: class formation, class expansion and finding common attributes (with learning tasks using the process of generalisation), identifying dissonant items (by discrimination), four-, six- or nine-fold schemes (by cross-classification), series completion, order series and simple analogies (by recognising relationships), disrupted series (by differentiating relationships) and matrices with complex analogies (using system formation). The instructions for the tasks were provided in written form on-screen. The language of instruction was simplified Arabic, and Eastern Arabic numerals were used. The 120 tasks (20 per operation) were divided into five sessions with 24 training tasks each. Figure 6.16 shows sample tasks for each of the operations. The training is based on a Hungarian training IR programme (Pásztor, 2016).

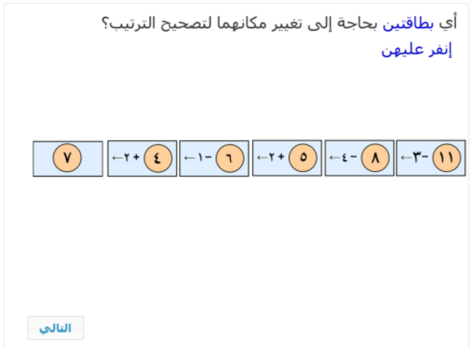
The effectiveness of the intervention programme was measured with the same internationally widely used computer-based test of inductive reasoning, the online assessment eDia system (see Molnár & Csapó, 2019a). A distinct context with different task types (figural series, figural analogies, number series, and number analogies) was used to avoid the near transfer effect of the training programme (see Molnár et al., 2013; Pásztor et al., 2018; Wu & Molnár, 2018a, 2018b). The original inductive reasoning test was modified as explained earlier (see Figure 6.3, 6.6., 6.17). Since Palestinian schoolbooks use the Eastern Arabic numeral system, items containing numbers were modified accordingly: Figure 6.17 shows the original and the adapted version of the same item from the test. The test consisted of 44 items in total, containing no interactive elements. The reliability index for the whole test was Cronbach's $\alpha = .812$ on the pre-test and Cronbach's $\alpha = .912$ on the post-test.

At the end of the training tasks, participants in the experimental group received some questions, in form of a short questionnaire (see Table 6.8 and see appendix D) regarding their views about the training program. We have highlighted the answers with three different colours (green: positive, yellow: neutral, red: negative) to make it easier to conclude.

<table>
<tr><td></td><td>Attributes of objects</td><td>Relations between objects</td></tr>
<tr><td>Similarities</td><td>Generalisation


Form three groups. Drag the items belonging to the same group onto the shelves.</td><td>Recognising relations


What is the order of the numbers?
Drag the next number to the question mark.</td></tr>
<tr><td>Differences</td><td>Discrimination


Which of the followings does not fit?
Click on it.</td><td>Discriminating relations


Which two cards need to be inverted for the correct sequence?
Click on them.</td></tr>
<tr><td>Similarities and differences</td><td>Cross-classification


The books on the shelves have been arranged according to a certain rule.
What can the rule be? Based on the rule, which shelf would you put the separate book on?
Click on it.</td><td>System formation


Observe what happens in the machines.
Drag the correct bags to the question mark.</td></tr>
</table>

Figure 6.16. Examples of training tasks

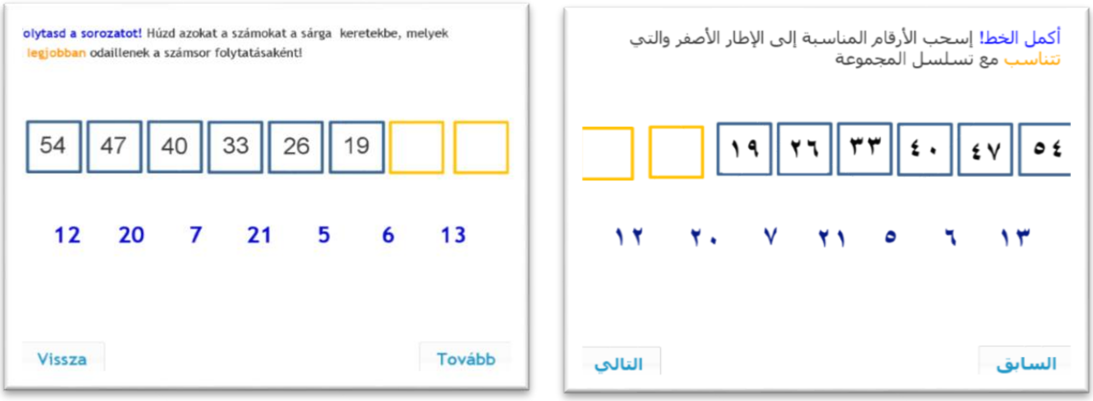


Figure 6.17. The original (Hungarian) and the adapted (Arabic) version of the same test item

Table 6.8. The questionnaire for the pupils in the experimental group (green: positive, yellow: neutral, red: negative)

No	Question	1	2	3	4	5
1	Did you like the game?	I liked it very much	It was okay	Liked and disliked as well.	I didn't like it	I didn't like it at all
2	How did you feel while playing?	Very well	Good	Sometimes well, sometimes I felt bad	Bad	Very bad
3	Have you ever got bored while playing?	Permanently	Often	Sometimes	Rarely	Never
4	If you had the chance, would you like playing such a game again?	Surely	Probably	Perhaps	Probably not	Surely not
5	If you had the chance, would you play at home with such a game?	Surely	Probably	Perhaps	Probably not	Surely not
6	How difficult were the tasks?	They were quite difficult.	They were difficult.	Sometimes they were difficult, sometimes they were not.	They were not difficult.	They were not difficult at all.
7	How much did the text of the task help you when you did not know the answer at once?	It helped quite a lot.	It helped a lot.	Sometimes it helped, sometimes it didn't.	It didn't help.	It didn't help at all.
8	Are you allowed to use your computer at home?	yes	no			

Pupils needed to choose the frequency 5-scale (Likert-scale) in seven out of eight questions. The results of these questions reflected pupils' views about the training program in general. The questionnaire also interpreted pupils' emotions during the training, for example, it investigated if they liked the training, the way they felt during the training and things related to the difficulty level of the training tasks and so on. Generally, the positive answers represented pupils' satisfaction and engagement in the training program.

6.4.2.3 Design and procedures

Study 4 used a quasi-experimental design, with both the experimental and control groups assessed at two different points in time, once before the intervention and once after the training. Pupils were randomly assigned to the experimental and control groups.

The computer-based IR training was administered via the eLea online training platform (Molnár, Pásztor, & Csapó, 2019), and the pre-test and post-test were delivered through the eDia online assessment platform (Csapó & Molnár, 2019; Molnár & Csapó, 2019a). The test and the training were administered on desktop computers in the computer labs at the participating schools. The training lasted six weeks and took place during regular school hours, after having received official approval from the directorate of education in the region (Bethlehem district) and having consulted with the schools' principals. Each session took approximately one school lesson, that is, forty minutes. Classroom teachers supervised the training.

There was a team of teachers who supervised the training sessions. One week before the training, the implementers received a short (three-hours-long) on-site group training held by the researchers. The aim of this meeting was to provide full information about the theory, platform and process of the training: (1) how thinking skills and abilities, especially inductive reasoning, develop during the age range of schooling; (2) what the characteristics of inductive reasoning targeted in the training were; (3) what the structure of the training looked like; (4) how much time was required for the project; (5) what the training tasks looked like – with examples; and (6) how the online training (eLea) and test (eDia) platforms could be used; that is, the teachers were taught how to run the system. During the training, it was agreed how to behave during the training sessions – when to interact with the children and what type of support was allowed during the training if the pupils asked for assistance – so as to make sure that all the pupils had the same opportunity to receive assistance. Teachers were allowed to resolve any technical issues and answer questions raised by the pupils but were not allowed to provide solutions for or make any reference to the right answer on any of the tasks. Teachers also learned how to track pupils' progress during the training.

Each pupil got a password and logged into the system with the help of the class teacher, as well as the ICT teacher, who gave a hand if any technical issues arose and facilitated the process of dealing with technical issues during the training. At the beginning of the training, pupils were provided with instructions on how to use the programme and how the training would run. The instructions included the following: in each session, a yellow bar at the top of the screen indicates how far along the pupils are on the test. When they finish a training task, they click on the "next" button to check the answer, or, if the answer is correct, proceed to the next task. Note that after each task, immediate feedback is provided on whether the answer is right or wrong. If the answer is right, the pupil receives the number of stones representing how many stones (right answers) he or she has collected so far. If the answer is wrong, the pupil receives additional supporting information and has another chance to complete the training task again. If the pupil fails a second time, he or she has a third and final chance to do the same task with much more help. When the pupil reaches the end of a session, he or she receives information about his or her achievement on the training.

Since the eDia assessment system (for the pre- and post-test) and the eLea training platform (for the intervention) logged, recorded, and scored the pupils' answers, immediate feedback on both the test and task levels was possible. Both the pre- and post-test took about one hour to complete, and similarly to the training sessions, classroom teachers supervised the collection of data, which happened in the ICT rooms of the participating schools.

Beyond the analyses using observed variables (e.g. ANOVA analyses), which have several limitations, we also used second-order multiple group latent curve modelling in the structural equation modelling (SEM) framework using latent variables based on Alessandri et al. (2017). Analyses underlying classic parametric tests operate according to the twin assumptions that the measurement structure of the construct under investigation is invariant across groups and/or time and that the data being analysed are normally distributed with equal population variances. As the latter one was not completely confirmed by the results, we decided to run SEM analyses to confirm the results obtained on the manifest level.

There were only two points in time available in the analyses; that is, it was possible to estimate two latent curve models: a no-change model (or strict stability model, see Alessandri et al., 2017) and a latent change model. In the no-change model, both the mean and variance of the second-order intercept factor were freely estimated across groups. This included only a second-order intercept factor representing the pupils' initial skill level. In the latent change model, a slope growth factor was estimated.

Generally, two waves of data are insufficient to estimate latent change models. There is, however, a way to do this if the latent change model is over-identified, that is, if at least two observed indicators for the construct of inductive reasoning are available at each time point, for example, if there are two scales assessing the same construct in the study. Although the construct of inductive reasoning was assessed by one scale (with the test composed of several items), it was possible to partition the items composing the scale into two parcels that could be treated as parallel forms, based on Steyer et al. (1997). We followed the procedure described in Little et al. (2002) in that we ran three alternative models: (1) a no-change model in both groups (experimental and control), (2) a latent change model for the experimental and a no-change model for the control group, and (3) a latent change model for both of the groups. We compared the model fit indices CFI (Comparative Fit Index) and TLI (Tucker–Lewis Index) with the RMSEA (Root Mean Square Error of Approximation) and associated 90% confidence intervals, as well as the changes in fit indices between the different models. We accepted CFI and TLI values > 0.90, RMSEA values < 0.08 (see Kline, 2016).

6.4.3 Results

6.4.3.1 The effectiveness of the online IR training using mathematical content at the ages of 9 to 11

No significant difference was found between the performance of the experimental group and that of the control group prior to the experiment (M_{cont} = 33.4, SD_{cont} = 15.0; M_{exp} = 35.0, SD_{exp} = 13.5, t = 1.3, p = .18). While no significant development could be detected in the case of the control group (M_{cont} = 34.0, SD_{cont} = 14.1), the experimental group significantly outperformed the control group by more than one standard deviation (M_{exp} = 58.6, SD_{exp} = 14.5, t = 13.1, p < .001).

Using Cohen's (1988) convention for describing the magnitude of effect size, we found a clear large effect $d = 1.71$. This effect size is as high as that published in previous literature (Klauer & Phye, 2008; Molnár, 2006) investigating face-to-face training programmes in a non-academic context using a pre- and post-test developed in accordance with the Klauer model. Therefore, placing our IR training programme in an international context independent of its delivery media, we may draw some favourable conclusions. The result indicates that it can be employed in mathematics lessons, improving pupils' inductive reasoning skills effectively, with development detected even on IR tests not devised in accordance with the Klauer model to avoid the near transfer effect of the training. (RQ19) (RQ20)

6.4.3.2 The changes in performance as regards pupils' original level of inductive reasoning skills

Figure 6.18 presents the item/person map, the numbers to the right side are the distribution of the items where at the top are the most difficult and at the bottom are the easy ones. The distribution of the items according to their difficulty level was good (see Figure 6.18). There were easy, medium, and hard items (even there were more hard items compared to the medium and easy ones), thus the test was suitable for measuring the target populations' inductive reasoning skills.

In the case of the control group, the distribution curve for both the pre- and post-test (see Figure 6.19) is inclined to the left. The two curves coincide approximately, indicating no "spontaneous" development in pupils' IR skills in the time period of normal school learning given. The post-test distribution curve for the experimental group, skewed to the right, reflects an immense improvement at each level of IR in the experimental group. Based on the group-level distribution curves, we may hypothesise that each member of the experimental group improved his or her performance significantly as a result of the training.

The group-level results above are supported by the pupil-level analyses, illustrated in Figure 6.20, where the performances in the pre- and post-tests are projected onto each other. The abscissa shows the achievement obtained in the first wave of data collection and the ordinate displays this from the second wave, where each dot represents a pupil. If a pupils' dot falls on the mean line or between the two broken lines (representing one standard deviation), he or she performed identically in the two cases. If the dot is positioned above the broken line, it means that the particular pupil showed significant development from the pre-test to the post-test. Finally, if the dot lies below the broken line, it represents a significantly worse performance in the second round of data collection.

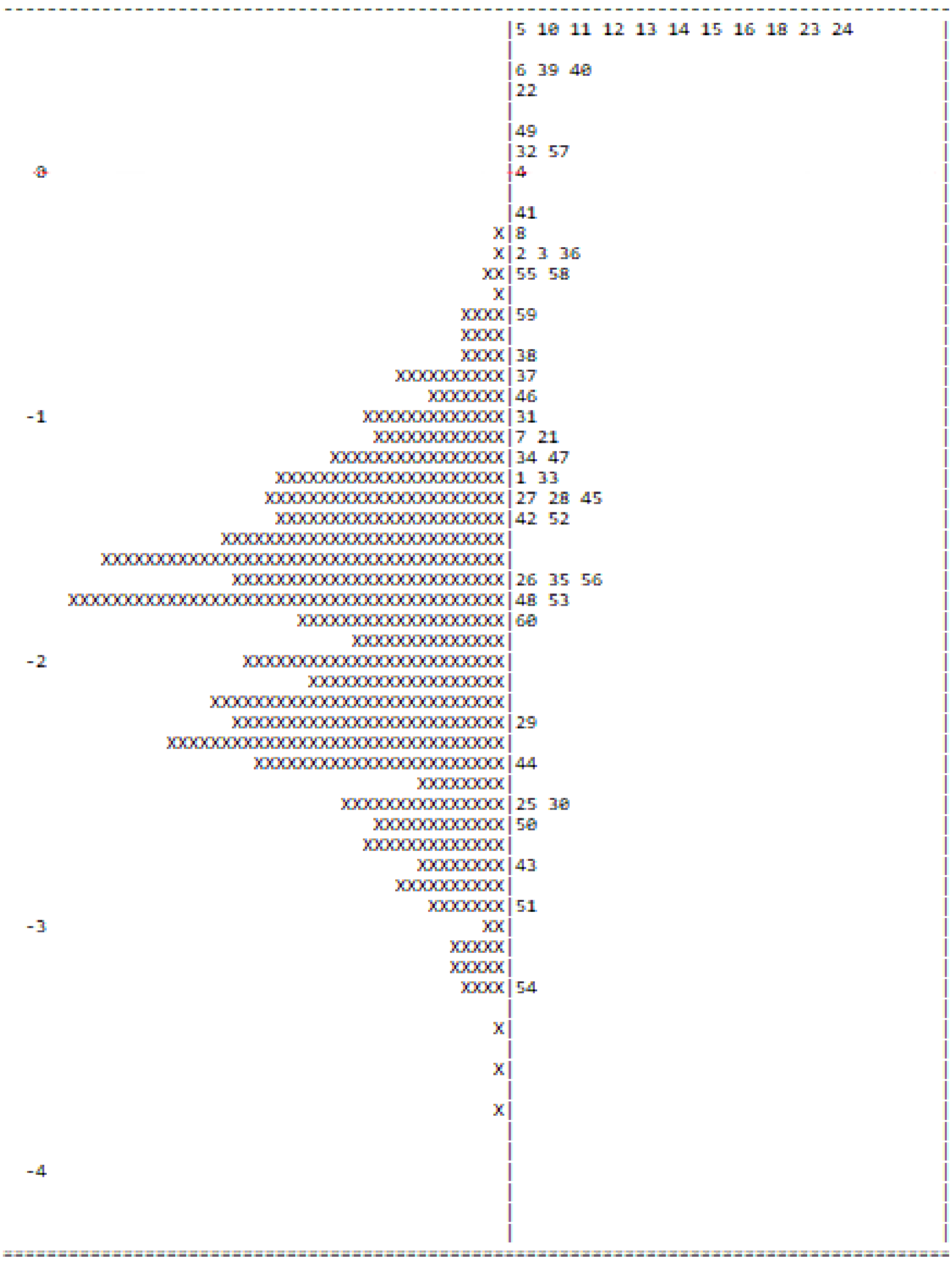


Figure 6.18. The one-dimensional item/person map (each x represents 0.2 pupils) of the experimental group in the pretest

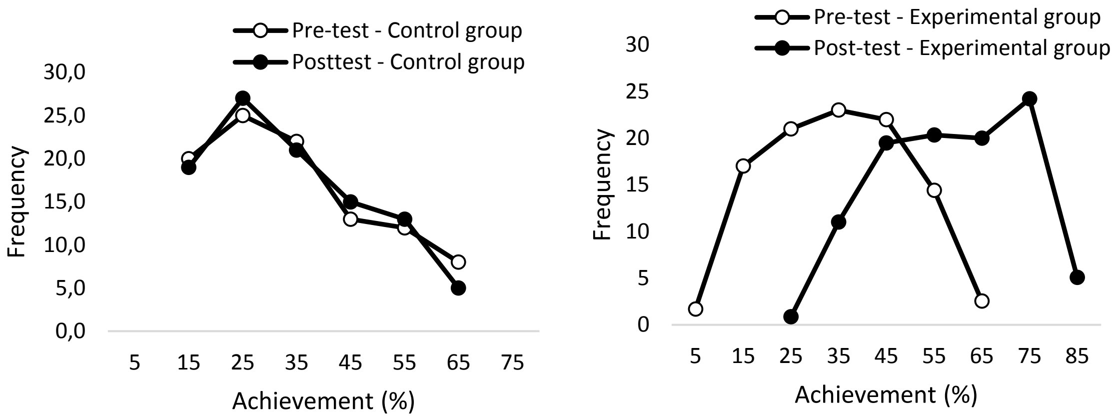


Figure 6.19. Distribution curves for the control and experimental groups in the pre- and post-test

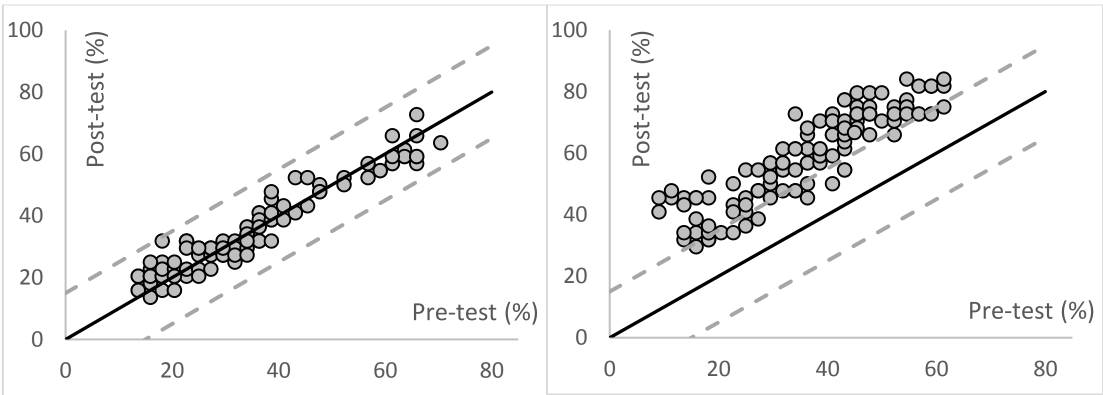


Figure 6.20. Pupil-level changes in achievement from pre-test to post-test in both the control and experimental groups

In the case of the control group (graph on the left), the dots are distributed around the mean line between the broken lines; that is, all the pupils in the control group performed significantly at the same level, scoring quite similarly on the two tests. A completely different tendency is displayed on the right-hand graph, showing the performance of the experimental group before and after the training. Almost all the dots are located above the broken line. That is, there is no pupil whose performance dropped significantly from pre-test to post-test in the time frame given, while almost all the experimental pupils' IR levels improved by more than one standard deviation as a result of the training. To sum up, the training led to a significant (p < .001) improvement in inductive reasoning for the experimental group. (RQ21)

6.4.3.3 The influential factors of pupils' gender, school achievement and socio-economic background on the magnitude of their development in inductive reasoning

The paired t-test results indicated no gender-level differences in the training effect; that is, there were no significant changes detected in the control group, while statistically significant differences were found in the experimental group in both subgroups. Based on the independent t-test, there were no significant gender-level differences in the pre-test or post-test results (see Table 6.9). Therefore, the training proved to be gender-fair.

The mother's educational background correlated strongly with the pupil's IR skill level at both time points ($r\_pre = .667$, $r\_post = .555$, $p < .001$) and in both groups (control group: $r\_pre = .718$, $r\_post = .741$; experimental group: $r\_pre = .626$, $r\_post = .642$; $p < .001$). This was also confirmed by the ANOVA analyses. Based on the Tukey B analyses, there were four groups whose achievement was significantly different at both time points: children whose mother (1) was undereducated or educated at the primary level (M_pre = 20.9/24.2; M_post = 32.4/36.1), (2) had a high-school certificate (M_pre = 35.6, M_post = 46.60, (3) had a diploma (A levels) (M_pre = 44.5, M_post = 55.6), or (4) had a BA/BSc or MA/MSc degree (M_pre = 49.5/52.3, M_post = 65.6/77.3). That is, pupils' IR skill level was generally higher if their mother had had more education, and vice versa.

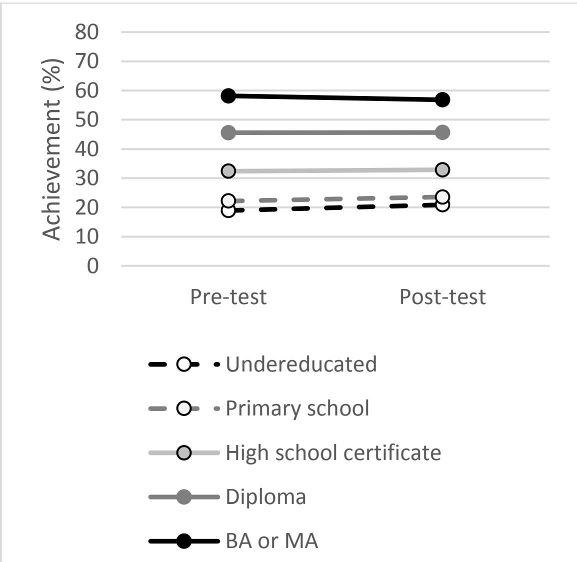
Table 6.9. Within and between gender-level differences on the pre- and post-test

<table>
<tr><th colspan="2">Group</th><th>N</th><th>r</th><th>Diff. in mean</th><th>SD</th><th>t_pre_post (p)</th><th>t_pre (p)</th><th>t_post (p)</th></tr>
<tr><td rowspan="2">Control group</td><td>Male</td><td>62</td><td>.964</td><td>.84</td><td>4.4</td><td>1.52 (.14)</td><td rowspan="2">-.97 (.34)</td><td rowspan="2">-.95 (.34)</td></tr>
<tr><td>Female</td><td>56</td><td>.958</td><td>.52</td><td>4.5</td><td>.87 (.39)</td></tr>
<tr><td rowspan="2">Experi-mental group</td><td>Male</td><td>65</td><td>.887</td><td>24.05</td><td>7.4</td><td>26.11 (.001)</td><td rowspan="2">-.65 (.52)</td><td rowspan="2">.54 (.59)</td></tr>
<tr><td>Female</td><td>53</td><td>.924</td><td>21.01</td><td>5.0</td><td>30.71 (.001)</td></tr>
</table>

Note. r: correlation

Figure 6.21 compares the performance differences of pupils with mothers with different educational backgrounds. In the case of the control group (graph on the left), no change can be detected in either of the groups. Meanwhile, in the case of the experimental group (graph on the right), the trajectories representing the effects of the training are parallel, thus indicating similar training effects across all groups. That is, none of the four subgroups was favoured as regards the efficacy of the training in the experimental group.

Finally, pupils were divided into three subgroups based on their achievement in school. A strong relation was detected between school achievement and IR skill level (pre-test: $M\_advanced = 53.78$, $M\_average = 33.99$; $M\_low = 19.67$, $F = 460.43$, $p < .001$; $r = .889$, $p < .001$; post-test: $M\_advanced = 67.06$, $M\_average = 44.89$; $M\_low = 31.07$, $F = 136.16$, $p < .001$; $r = .726$, $p < .001$). Figure 6.22 illustrates the trajectories for IR in the control and experimental groups based on pupils' school achievement. The developmental curves are parallel for the experimental group; that is, the training had the same effect on all the pupils, independent of their school achievement and initial IR skill level. (RQ22)



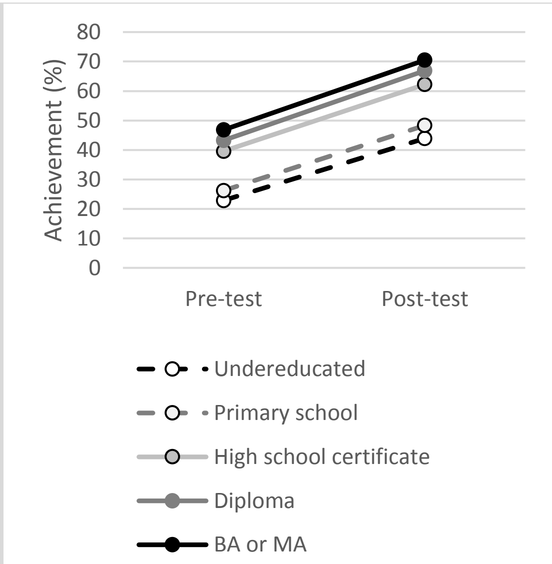


Figure 6.21. Trajectories for inductive reasoning skills in the control group (graph on the left) and the experimental group (graph on the right) according to the mother's educational attainment

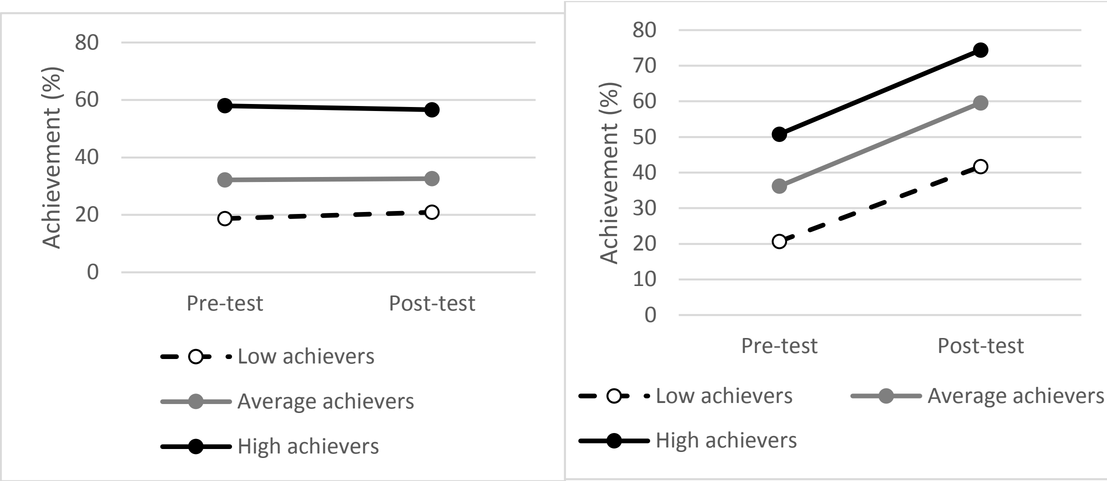


Figure 6.22. Trajectories for inductive reasoning skills in the control group (graph on the left) and the experimental group (graph on the right) according to the pupils' school achievement

6.4.3.4 Evaluating the effect of the intervention programme in the latent curve modelling framework

First, we tested a measurement model for inductive reasoning with all the indicators combined under one general factor. We used the preferred estimator for categorical variables, Weighted Least Squares Mean and Variance adjusted (WLSMV; Muthén & Muthén, 2010). The measurement model based on the pre-test results showed a good fit (χ^2 = 974.9; *df* = 902; *p* < .05; *CFI* = .940; *TLI* = .937; *RMSEA* = .019 (*CI*: .003, .027)).

We created two parallel forms of the inductive reasoning scale based on the factor loading values. The Cronbach's alphas were good (≥ 0.79), and correlations were above .89. Table 6.10 shows the fit indices for the three models. As hypothesised, the mixed second model was the best fitting model (a no-change model for the control group and a latent change model for the experimental group); however, the RMSEA value was still higher than acceptable for a good model fit. All the fit indices for the other two models fall below accepted values.

The results also confirm that there was no significant variability among pupils in responding to the intervention programme, as indicated by the non-significant variance of the latent slope. None of the pupils were more sensitive or responsive to the intervention delivered. (RQ23)

Table 6.10. Goodness-of-fit indices for the models tested

Model	χ^2 *(df)*	*CFI*	*TLI*	*RMSEA* [90% *CI*]
No-change model for both groups	615.8 (12)	.602	.602	.653 [.610, .697]
No-change model for the control group and latent change model for the experimental group	47.7 (11)	.974	.961	.205 [.151, .263]
Latent change model for both groups	312.6 (10)	.801	.761	.506 [.459, .555]

6.4.3.5 Experimental group participants' views about the training program

The results of the questionnaire in general were positive (see Figure 6.23), pupils' responses were in favour of the training. Most of the pupils (92.9%) said they liked the game and 84.9% of them felt good while playing the game. Half of the pupils said that they sometimes felt bored while doing the tasks and 37.3% of them did not get bored. Pupils were positive (77.5%) when responding to play the game once again if they got the chance to do so, and they were also encouraged (74.5%) to play such a game even at home. About two-thirds (61.4%) of the pupils felt that the tasks were difficult, but a lot of them (87.4%) said that when they did not know the answer, the text of the task helped them. Finally, the majority of the pupils (95.4%) said they do not have any problem using the computer at their home. (RQ24)

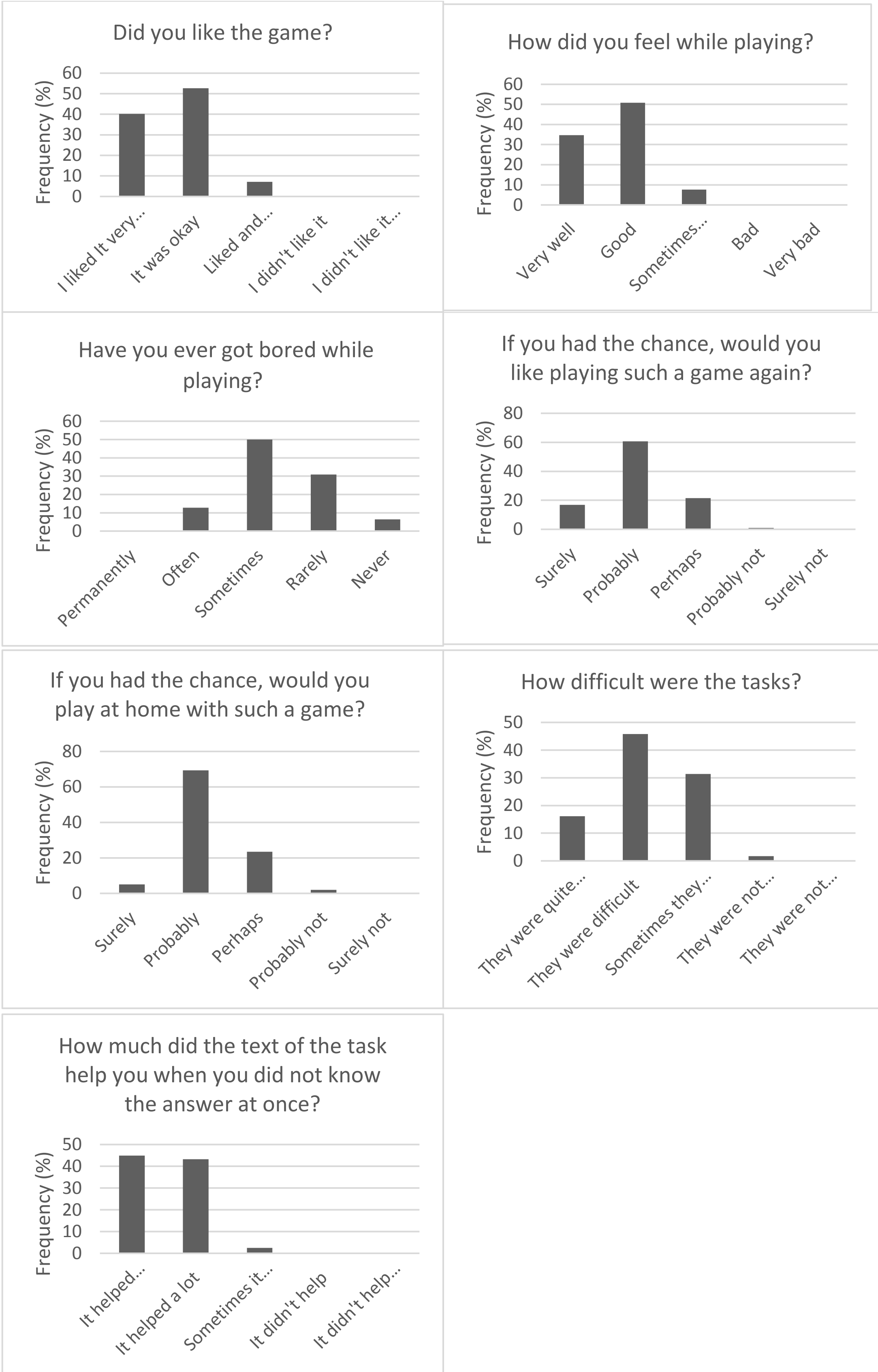


Figure 6.23. The views of the experimental group's pupils on the training program

6.4.4 Discussion

This study presented a computer-based training programme in inductive reasoning for 9–11-year-old pupils in an educational context and addressed the direct result of the evaluation study. The online intervention programme was developed according to Klauer's model and the "Cognitive training for children" concept of inductive reasoning; it thus consists of 120 playful problems, but in an online environment. All the problems were embedded in mathematical content to make it possible to integrate and deliver the training during normal school hours as part of the mathematics lesson.

The study used a quasi-experimental design. The developmental level of the control and experimental groups did not differ prior to the experiment, so the control group met the requirements for a control group. As a result of the online training, the inductive reasoning skills of the experimental group showed significant improvement, while pupils' IR skills remained at the same level in the control group. This result confirms previous research results as regards (1) the usability and effectiveness of technology-based trainings at the school level (see Molnár & Pásztor, 2015a; Mousa & Molnár, 2018, 2019a, 2019b); (2) the possibility of explicit fostering of inductive reasoning during normal school hours (Adams, 1989; de Koning et al., 2002; Molnár, 2011; Nisbet, 1993; Resnick & Klopfer, 1989), and (3) the lack of explicit fostering of inductive reasoning in school (de Konig, 2000; Molnár, 2011). Thus, we confirmed that Klauer's concept of cognitive training for children can be applied not only face-to-face, but in an online environment as well, and not only in an everyday context, but also in the context of a school subject, such as mathematics.

The effect size of the training was exceptionally good not only in an Arabic context, but internationally as well. It proved to be unrelated to gender, to the mother's educational attainment (which is known to be a good socio-economic indicator (Wegerif et al., 2017) – indeed, the mother plays the major role in educating children in the home in Palestine), to school achievement, and to the original IR level of the pupils; that is, it had a similar effect on boys and girls, on pupils whose mothers' educational attainment afforded them different kinds of stimulation and instruction at home, on pupils with low, average or high school achievement, and on pupils with a low or high starting level in IR at the ages of 9–11. This result was confirmed by the structural equation modelling analyses, where we used three models: three different combinations of no-change and latent change models in both the experimental and control groups. The best fitting trajectory (a no-change model in the control group and a latent change model in the experimental group) confirmed the results on the manifest level. The non-significant variance of the latent slope indicated and proved previous results of no significant variability among pupils in responding to the intervention programme. Generally, none of the pupils were more sensitive to the intervention delivered.

The vast majority of the experimental pupils answered the questionnaire at the end of the training program so it can be concluded that answering the questions was taken seriously. Looking through the answers they present pupils' satisfactions about the training, which proved that placing tasks into game-based mode can evoke their engagement in learning (see Yang & Chang, 2013).

To sum up, the results suggest that this intervention programme, specific to the online domain, was developed successfully. It improves 9–11-year-old pupils' inductive reasoning in a playful way. The findings suggest that inductive reasoning skills can be developed significantly and effectively between the ages of 9 and 11, not only in a traditional face-to-face environment, but also in a computer-based one. Furthermore, an online test of inductive reasoning was constructed in Arabic as part of the programme package. Given its structural validity and reliability, it can be used effectively in assessing the developmental level of primary school pupils' inductive reasoning skills even independent of the training programme.

7 CONCLUSIONS, RECOMMENDATIONS, IMPLICATIONS, AND LIMITATIONS

This research provided insights about the interests of the education systems worldwide in the 21st century regarding using modern educational assessment tools in the education systems. Study 1, 2 and 3 investigated the visibility and the applicability of computer-based assessment in the Palestinian schools when it comes to early age pupils. We wanted to see if the education system in Palestine develops pupils thinking skills or not. Palestine is not part of any international assessment programs like PISA or TIMSS. The aim was to make sure of pupils' ability of using basic mouse skills which is needed to use the online assessment, as the items require these abilities to answer the tasks. Then we wanted to discover the developmental level of thinking skills for the pupils in their early school age before implementing the enhancement program. International assessments indicated that there is a clear gap between boys and girls in the international level, so we wanted to know the case of Palestinian pupils by including analysis of gender differences. The studies also focused on the influence of background variables since they play an important role in pupils' performance.

Study 1 aimed at initiating computer-based assessment in Palestinian elementary schools. Technology-based assessment provides schools and teachers with a user-friendly instrument that can help in administering the development of thinking skills for pupils (Pásztor et al., 2015). We have noticed clearly the advantages of using the modern educational technologies while doing the data collection in the schools, however, there are some downsides regarding the infrastructure of the schools in general, and the ICT laboratories to be used for everyday school practise. It will be discussed in the limitation section.

Besides, the psychometric analyses of the mouse skills and the IR tests proved they were reliable for assessing elementary school pupils' abilities. However, a bigger sample size was needed to confirm the results and make it generalizable. So far, we can say that hypotheses (H_1-H_3) were confirmed since there might be sometimes differences in the results when it comes to the sample size (see Mousa & Molnár, 2018, 2019a).

Literature presents that mothers' educational and occupational background might affect the children's outcomes i.e. achievement (see Csapó, 2010; Nikolov & Csapó, 2018; Pásztor, 2016; Nikolov & Csapó, 2009), therefore, we wanted to know if there were any relations between the background variables and the achievement of the pupils. In this study (Study 1), we could not confirm our hypothesis (H_4) since surprisingly only mothers' occupation correlated with the children's IR achievement but mothers' educational background did not.

Generally, in the Palestinian schools for the elementary level, it is not expected to find differences between boys and girls, while when it comes to later school grades, the gap between the two genders start to appear. It clearly can be seen from statistics that girls' performance in general is better than the boys' performance according to the Palestinian Central Bureau for Statistics (PCBS, 2013). In the literature, most of the assessments and the experimental studies showed no difference between the two genders, only when it comes to deeper analysis (see Weaver & Raptis, 2001; Clariana & Wallace, 2002; Hotulainen, et al., 2016; Molnár, 2011; Pásztor, 2016; Molnár & Csapó, 2019b; Csapó, Lőrincz, & Molnár, 2012). Therefore, we expected to find no gender differences between boys and girls for this early age in the IR test, surprisingly, there were gender differences and girls performed better than boys. Again, we can say here that a bigger sample size is going to give a more precise idea since what made the difference was detected in grade 2, so our hypothesis (H_5) was not confirmed.

The results revealed the fact that IR skills develop with age (see also Pásztor, 2016, Csapó, Molnár, & Kinyó, 2008; Molnár, Greiff, & Csapó, 2013), as pupils in the upper grade performed better in the test in comparison with the pupils in the lower grade, thus our hypothesis (H_6) was confirmed. We also compared the school achievement with the IR test results and we found that the performance of the pupils in the IR test correlated with their school achievement and that means pupils' achievement at their school influence their results which proves that inductive reasoning skills are available in their curriculum and their skills are being enhanced by it, and that confirms our hypothesis (H_7).

These results made it necessary to move forward to the next one, Study 2. The aim of Study 2 was to pilot a computer-based assessment test of inductive reasoning skills among second-, third- and fourth-graders. It also aimed to discover background factors like grade and mothers' level of education influencing the applicability of CBA with Palestinian pupils and test gender differences regarding inductive reasoning. The online IR test for assessment proved to be applicable and reliable for the whole test in the three different grades. The Cronbach's alpha results for both subtests (figural series and figural analogy) were good (α >.80), that means the format of the task does not influence the performance of the pupils, hence our hypotheses (H_8 & H_9) has been confirmed. Therefore, we took the next step of developing the test by adding more difficult items and some numerical items to be used for the enhancement program.

According to the findings of the current study, pupils' inductive reasoning skills are in general being enhanced in their early school age with only being taught their school curriculum. In other words, the pupils were not subjected to any enhancement training from outside to affect their development, therefore, the Palestinian school curriculum contained inductive reasoning skills in it in the second, third and fourth grades (see also Mousa & Molnár, 2019a), here we confirmed our hypothesis (H_{10}).

This study confirmed our hypothesis (H_{11}) about the gender differences. After using a larger sample size, in the pilot study detected gender differences disappeared, that is, the gender differences detected in grade 2 (Mousa & Molnár, 2019a) were caused by the small sample size.

Our study supports the literature (see Nikolov & Csapó, 2018; Csapó, 2010) with the idea that background factors influence pupils' performance. In our first study (Mousa & Molnár, 2019a) we could not prove it. We claimed that it might be related to the small sample size of that study. For this study (Study 2), we proved that the mothers' background variables influence pupil's performance, thus we confirmed our hypothesis (H_{12}). Another important question might be asked, whether the teaching and learning process in Palestine develop pupils' IR skills. In this study, we also confirmed our hypothesis (H_{13}) regarding the strong correlation between the school achievement and the performance in the test, that means pupils' thinking skills develop during the classroom process of teaching and learning.

The aim of Study 3 was to explore the possibility of using computerized tests with pupils in an early school age (fourth and fifth grades) by assessing IR skills. It also aimed to collect data measuring the pupils' inductive reasoning skills using an online test. We also wanted to find out if background factors i.e. grade and mothers' level of education influence the applicability of CBA in case of Palestinian school children and test gender differences regarding inductive reasoning skills. We confirmed that the online inductive reasoning test was reliable, and the psychometric properties of the test were acceptable. The results of the developed IR test, which includes more difficult items and items with numerical values proved to be reliable with Cronbach's alpha (α = .807), it also means that the results of the study are generalisable (see Molnár & Csapó, 2019c; Pásztor, 2014; de Koning et al., 2002; Klauer & Phye, 2008; Perret, 2015) and here hypothesis (H_{14}) was confirmed. We also confirmed our hypothesis (H_{15}), as it was the case in the previous studies (Mousa & Molnár, 2018, 2019a), that computer-based assessment is applicable with the Palestinian school pupils in their age between 9-10 years old and also in mathematics.

The developed mathematical context which has been added to the test proved to be much harder when it came to numbers items in comparison with figures items. Fifth grade pupils achieved significantly higher than the fourth-graders, that confirms hypothesis (H_{16}), which means pupils' inductive reasoning skills enhancement is sensitive for them at this school age even without having any training. This answers the question raised previously regarding the teaching and learning process in Palestine and the development of pupils' IR skills which proved, that the IR skills can be enhanced at early school age with just being taught by the textbooks of the school curriculum and without being subjected to any training from outside that affects the development (see also Mousa & Molnár, 2018; 2019a).

We found no significant gender-level differences in the achievement on the inductive reasoning test and that confirms hypothesis (H_{17}). Both male and female pupils achieved at the same level in the same grade and this study proved the results we got in the previous study (Mousa & Molnár, 2018) when we went for a bigger sample size. However, when it came to the behaviour during the test, some differences between the two genders could be detected and that goes along with the literature' findings (Molnár & Csapó, 2019b; Csapó, Lőrincz, & Molnár, 2012), that differences between the two genders are detectable with deeper analysis.

A bigger sample size proved what we could not do with a small sample size. The background factors' influence on pupils' performance has been confirmed in this study, as in the literature (see Csapó, 2010; Nikolov & Csapó, 2018; Pásztor, 2016; Nikolov & Csapó, 2009). The correlation between the mother's educational background and the pupils' achievement in the IR test proved to be strong, that means the higher the mother' education is, the higher the achievement obtained. It is the same case for the mother's occupational background which strongly influenced the achievement of the pupils, this conformed hypothesis (H_{18}). We confirmed hypothesis (H_{19}) that pupils with higher achievements at their schools performed better in the inductive reasoning test. These results go along with the results we got in the previous studies (see also Mousa & Molnár, 2018, 2019a, 2019b) regarding the school curriculum in Palestine and IR skills.

The last study (Study 4) aimed on the one hand to examine the applicability and effectiveness of an online inductive reasoning training programme based on Klauer's "Cognitive training for children" concept in an Arabic educational context for the group age of 9–11 years old (fourth and fifth grades). On the other hand, it investigated the effectiveness of the intervention programme on different groups of pupils regarding their levels of IR, socio-economic factors and gender differences. We confirmed our hypothesis (H_{20}) that the developed training programme can improve pupils' inductive reasoning skills effectively. The results go along with the literature in the point that the improvement in pupils' performance can be effective even in an online environment using mathematical task content (see Csapó & Szendrei, 2011; Csapó & Csépe, 2012; Csapó & Szabó, 2012; de Koning et al., 2002; Klauer & Phye, 2008; Perret, 2015; Pásztor, 2014). The online training programme which was based on Klauer's training concept in inductive reasoning also proved to be effective in an Arabic educational context for the age range 9–11years old and that means it can be applied internationally, this confirmed our hypothesis (H_{21}). The results showed that the improvement affected each member of the training (H_{22}) and the effect size of our programme met the previous literature's (Klauer & Phye, 2008; Molnár, 2006) conclusions which put our training programme to a global level.

The training proved to be gender-fair regarding the differences between the male and female participants. There were no significant differences detected between the two genders in all tests and that confirmed our hypothesis (H_{23}) which we based on the results we achieved from the previous studies (Mousa & Molnár, 2018, 2019a, 2019b) and also regarding the findings from the literature (see Weaver & Raptis, 2001; Clariana & Wallace, 2002; Hotulainen et al., 2016; Molnár, 2011; Pásztor, 2016). The background variable played an important role in influencing the pupils' IR skills level (H_{23}). The higher the mother' education level was, the higher the IR skills level occurred. The level of IR skills decreased as the mother's education was lower, thus we supported the literature findings (see Csapó, 2010; Nikolov & Csapó, 2018; Pásztor, 2016; Nikolov & Csapó, 2009) and we agree that these background variables should be taken in consideration regarding their importance that might make difference in the achievement of some pupils.

The results of the study proved that the pupils have no significant variability in responding to the intervention programme. That means our intervention programme was not sensitive or responsive to the participating pupils, in this regard, the results confirmed our hypothesis (H_{24}). We were also able to confirm hypothesis (H_{25}) by being engaged into the training tasks in its game-based style (see Yang & Chang, 2013).

Due to the context of the study, the first piloting phase of a larger project, there were smaller sample sizes available for the analyses. A further limitation of the present study is that only pupils' mouse skills and not keyboarding skills have been tested to monitor pupils' basic ICT skills. This deficiency may be rectified by extending future investigations to mouse and keyboarding skills and larger sample size.

The training programme was organised for a limited timeframe of six weeks regarding several reasons including the limited timeframe of my study in Hungary and the limited number of computers at schools' computer labs (15-20 devices), where it was so difficult to occupy the room for a long time without interrupting the school schedule, since the education office gives a permission to do such a study only in case it does not influence the schools' study-schedule. We suggest for further research to cover more subject areas and to be more intensive, so it can give a bigger picture.

The results of the analysis showed that the numerical items were more difficult than the figural ones. We do not have a clear answer to what the reason could be, but it might be related to the location of the items at the end of the test, where the pupils either may not have enough time for the last items as they spend more time on the first parts, the figural, meaning there was a lack of time or the pupils were weaker in numbers. Some studies pointed out that the item order in computer-administered test might affect performance on a specific item (see Clariana & Wallace, 2002). This needs more investigation and analysis to confirm. Therefore, it is suggested to research this point for clarification.

Regarding the training, the limitations of the study include the procedure of dividing the pupils into control and experimental groups, resulting in two groups with the same level of inductive reasoning skills, but different socio-economic background factors, e.g. mother's level of education. Further repetition is required to validate the results with a larger sample and groups with not only the same average level of inductive reasoning skills, but also other background factors.

To avoid the near-transfer effect of the training, different models and different types of tasks were used in developing the pre- and post-test tasks. While the training tasks were developed according to Klauer's model of inductive reasoning to activate generalisation, discrimination, recognising relations, discriminating between relations and cross-classification, the test consisted of tasks involving figural and numerical analogies and series. This resulted in a more valid assessment of the effect size of the training. It also ruled out the option of analysing the effect size of the training based on the pre- and post-test data on the dimensional level. The dimensional-level analyses require logfile analyses of the training sessions, which might provide more detailed information on the training tasks themselves and help to improve the effectiveness of the programme. This forms part of our future plans, but it is also among the limitations of the present analysis.

I certify that the content of this dissertation is my own work of production. This dissertation has not been submitted for any other degree previously or at any other educational institution.

ACKNOWLEDGMENT

Being abroad for a long period of time is not an easy decision. However, there are people who and things that make it possible to be away and reduce its difficulties. All thanks go to my supervisor who gave me this opportunity to do my PhD under her supervision and for her support and patience in guiding me through the whole process of studying. I would also like to thank the Stipendium Hungaricum scholarship for their financial support which made it possible to do the study work properly.

My deep appreciation goes to the stuff members of the Doctoral School of Education at the University of Szeged for their kind help and for my fellow researchers as well.

Big thanks to my lovely parents who encourage and support me in all my steps for a long time until this moment to continue my studies. I am so proud of you and all what you have given to me. If there is something I want to be, it will only be just reaching your perfection.

A special gratitude to my lovely wife Ujfalusi Beatrix who joined me in the middle of my study and shared her life with me. You make me feel Hungary like home. Being with you is the best decision I have ever taken in my life.

To all my friends in Hungary, I would like to thank you all for the happy time we spent together and for all your support you gave me on my wedding day. We came from different parts of this World, but we were able to make the distance smaller by creating the goodness of friendship. I am so lucky to have you all in this period of time and in this specific place.

8 PUBLICATIONS RELATED TO THE DISSERTATION

Mousa, M. & Molnár, G. (2020). Computer-Based Training in Math Improves Inductive Reasoning of 9- to 11-year-old Children. *Thinking Skills and Creativity*, 37. doi: 10.1016/j.tsc.2020.100687

Mousa, M. & Molnár, G. (2020, June 22-26). Improving Pupils' Inductive Reasoning by Computer-Based Training in Maths. SIG 1 + 4 Conference 2020 in Universidad de Cadiz, Spain, Cadiz, Spain. https://earli.org/SIG1andSIG4-Cadiz2020 (conference cancelled).

Mousa, M. & Molnár, G. (2019a). The feasibility of computer-based testing in Palestine among lower primary school students: Assessing mouse skills and inductive reasoning. *Journal of Studies in Education*, *9*(2), 1–16. doi:10.5296/jse.v9i2.14517

Mousa, M. & Molnár, G. (2019b). Applying computer-based testing in Palestine: Assessing fourth and fifth graders inductive reasoning. *Journal of Studies in Education*, *9*(3), 1–13. doi:10.5296/jse.v9i3.14799

Mousa, M. & Molnár, G. (2018). Introducing computer-based assessment among 2nd to 4th grade students in Palestine. *16th Conference on Educational Assessment (CEA)*, Szeged, Hungary, 26-28 April, 2018.

Mousa, M. & Molnár, G. (2018). Assessing the 2nd & 3rd Years pupils' inductive reasoning by introducing computer-based testing in Palestine. *9th Biennial Conference of EARLI SIG 1: Assessment and evaluation: Assessment & Learning Analytics*, Helsinki, Finland, 29-31 August, 2018.

Mousa, M. & Molnár, G. (2019). The efficacy of a computer-based cognitive training program to enhance 4th and 5th grade students' inductive reasoning. *17th Conference on Educational Assessment (CEA)*, Szeged, Hungary, 11-13 April, 2019.

Mousa, M. (2017). Technology based assessment and enhancement of thinking skills: A case study of the educational system development in Palestine. *International Humanities Studies, 4*(2), 22–33.

Mousa, M. (2017). The influence of inductive reasoning thinking skill on enhancing performance. *International Humanities Studies, 4*(3), 37–48.

9 REFERENCES

AbdulKader, K. (2014). Higher order thinking skills in Math curriculum books for higher primary level in Palestine from the point of view of the teachers. *The Islamic University's Journal for Educational and Psychological Studies, 22* (1), 54–31 (Arabic version).

Adams, M. J. (1989). Thinking skills curricula: Their promise and progress. *Educational Psychologist, 24*(1), 25–77.

Adams, P. C. (1998). Teaching and learning with SimCity 2000. *Journal of Geography, 97*(2), 47–55.

Adams, R. J. & Wu, M. L. (2002). *PISA 2000 technical report.* Paris: OECD.

Adey, P., & Csapó, B. (2012). Developing and assessing scientific reasoning. In B. Csapó & G. Szabó (Eds.), *Framework for diagnostic assessment of science* (pp. 17-53). Budapest: Nemzeti Tankönyvkiadó.

Adey, P., Csapó, B., Demetriou, A., Hautamaki, J., & Shayer, M. (2007). Can we be intelligent about intelligence? Why education needs the concept of plastic general ability. *Educational Research Review, 2*(2), 75–97.

Adey, P. & Shayer, M. (1994). *Really raising standards: Cognitive intervention and academic achievement*. London: Routledge.

Adey, P., Shayer, M., & Yates, C. (2001). *Thinking science: The curriculum materials of the CASE Project*, 3rd Edn. London: Nelson Thornes.

Adler, J. (2017). Mathematics in mathematics education. *South African Journal of Science, Volume 113*(3/4). doi: 10.17159/sajs.2017/a0201

Affouneh, S. (2014). Critical analysis of the educational system after the establishment of the Palestinian Authority. *An-Najah University Journal for Research – Humanities, 28*(2), 266–292 *(Arabic version).*

Aizikovitsh, E. & Amit, M. (2010). Evaluating an infusion approach to the teaching of critical thinking skills through mathematics. *Procedia-Social and Behavioral Sciences, 2*(1), 3818–3822.

Alessandri, G., Zuffianò, A., & Perinelli, E. (2017). Evaluating intervention programs with a pretest-posttest design: A structural equation modelling approach. *Front. Psychol, 8*(223). doi: 10.3389/fpsyg.2017.00223

Alhadath. (2016). The Ministry of Education announces digitizing educational program at a cost approximately 20-24 million dollars *(Arabic version).* Alhadath Alfilistini Newspaper. Retrieved August 8, 2020, from: http://www.alhadath.ps/article/33138/

Arslan, S. & Yıldız, C. (2010). Reflections from the experiences of 11th graders during the stages of mathematical thinking. *Education and Science, 35*(156), 17–31.

Asad khan, R. M., Iqbal, M., & Tasneem, S. (2015). The influence of parents' educational level on secondary school students' academic achievements in district Rajanpur. *Journal of Education and Practice, 6*(16), 76–79.

Aunio, P. & Räsänen, P. (2015). Core numerical skills for learning mathematics in children aged five to eight years – a working model for educators. *European Early Childhood Education Research Journal, 24*(5), 684–704.

Backwell, J. & Hamaker, T. (2004). Cognitive acceleration through technology education (CATE): Implications for teacher education. Paper presented at epiSTEME -1 International Conference to review research on Science, Technology and Mathematics Education, Goa, India, 2004.

Barbak, A. (2012). The level of using creative thinking by Islamic education teachers for ninth grade students at UNRWA schools in Gaza. *The Islamic University's Journal for Educational and Psychological Studies, 20*(1), 91–129.

Barkl, S., Porter, A., & Ginns, P. (2012). Cognitive training for children: Effects on inductive reasoning, deductive reasoning, and mathematics achievement in an Australian school setting. *Psychology in the Schools, 49*(9), 828-842.

Barnes, S. K. (2010). Using computer-based testing with young children. In *NERA Conference Proceedings 2010*. University of Connecticut. Retrieved from https://opencommons.uconn.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1010&context=nera_2010

Becker, J. (2004). *Computergestütztes Adaptives Testen (CAT) von Angst entwickelt auf der Grundlage der Item Response Theorie (IRT).* Unpublished PhD dissertation. Freie Universität, Berlin.

Becker, K. (2001). Teaching with games - The Minesweeper and Asteroids experience. *The Journal of Computing in Small Colleges, 17*(2), 22–32.

Bejar, I. I., Lawless, R. R., Morley, M. E., Wagner, M. E., Bennett, R. E., & Revuelta, J. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *Journal of Technology, Learning and Assessment, 2*(3). Retrived from: http://escholarship.bc.edu/ojs/index.php/jtla/article/view/1663

Beller, M. (2013). Technologies in large-scale assessments: New directions, challenges, and opportunities. In M. von Davier, E. Gonzalez, I. Kirsch, & K. Yamamoto (Eds.), *The role of international large-scale assessments: Perspectives from technology, economy, and educational research* (pp. 25–45). Dordrecht, The Netherlands: Springer.

Bennett, R. E. (2003). *Online assessment and the comparability of score meaning*. Princeton, NJ: Educational Testing Service.

Bennett, R. E. (2002). *Using electronic assessment to measure student performance. The state education standard*. Washington, DC: National State Boards of Education.

Bennett, R. E. (2001). How the Internet will help large-scale assessment reinvent itself. *Education Policy Analysis Archives, 9*(5). Retrieved from: http://epaa.asu.edu/epaa/v9n5.html

Betz, J. A. (1995). Computer games: Increase learning in an interactive multidisciplinary environment. *Journal of Educational Technology Systems, 24*(2), 195–205.

Bisanz, J., Bisanz, G., & Korpan, C. A. (1994). Inductive reasoning. In R. Sternberg (Ed.), *Thinking and problem solving* (pp. 181–213). San Diego, CA: Academic Press. doi: 10.1016/B978-0-08-057299-4.50012-8

Bottino, R., Ferlino, L., Ott, M. & Tavela, M. (2007). Developing strategic and reasoning abilities with computer games at primary school level. *Computers & Education, 49*(1), 1272–1289. doi: 10.1016/j.compedu.2006.02.003

Breiter, A., Groß, L. M., & Stauke, E. (2013). Computer-based large-scale assessments in Germany. In D. Passey, A. Breiter, & A. Visscher (Eds.), *Next generation of information technology in educational management* (pp. 41–54). Berlin, Heidelberg: Springer.

Bridgeman, B. (2010). Experiences from large-scale computer-based testing in the USA. In F. Scheuermann & J. Bjornsson (Eds.), *The transition to Computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 39–44). Brussels: European Communities.

Brom, C., Šisler, V., & Slavík, R. (2009). Implementing digital game-based learning in schools: augmented learning environment of 'Europe 2045'. *Multimedia Systems, 16*(1), 23–41.

Brooker, L., & Siraj-Blatchford, J. (2002). 'Click on Miaow!': How Children of Three and Four Years Experience the Nursery Computer. *Contemporary Issues in Early Childhood, 3*(2), 251-273. doi:10.2304/ciec.2002.3.2.7

Bruer, J. T. (1993). *Schools for thought: a science for learning in the classroom*. Cambridge: MIT Press/Bradford Books.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge: Cambridge University Press.

Carson, K., Gillon, G., & Boustead, T. (2011). Computer-administrated versus paper-based assessment of school-entry phonological awareness ability. *Asia Pacific Journal of Speech, Language and Hearing, 14*, 85–101. https://doi.org/10.1179/136132811805334876

Chen, S. & Huang, P. (2013). The comparisons of the influences of prior knowledge on two game-based learning systems. *Computers & Education, 68*(1), 177–186.

Cheng, C. & Su, C. (2012). A game-based learning system for improving student's learning effectiveness in system analysis course. *Procedia – Social and Behavioral Sciences, 31*(1), 669–675.

Cheng, P. & Holyoak, K. (1985). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 17*(1), 391–416. https://doi.org/10.1016/0010-0285(85)90014-3

Choi, S. W. & Tinkler, T. (2002, April). Evaluating comparability of paper-and-pencil and computer-based assessment in a K-12 setting. In annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Christakoudis, C., Androulakis, G. S., & Zagouras, C. (2011). Prepare items for large scale computer-based assessment: Case study for teachers' certification on basic computer skills. *Procedia-Social and Behavioral Sciences, 29*, 1189–1198.

Clariana, R. & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with test mode effect. *British Journal of Educational Technology, 33*(5), 593–602. https://doi.org/10.1111/1467-8535.00294

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, NJ: Lawrence Earlbaum.

Connolly, T., Boyle, E., Macarthur, E., Hainey, T., & Boyle, J. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*(2), 661–686.

Corti, K. (2006). Games-based learning: a serious business application. PIXEL earning Limited. Retrieved from: https://www.cs.auckland.ac.nz/courses/compsci777s2c/lectures/Ian/serious%20games%20business%20applications.pdf

Csapó, B. (2010). Research on development in educational contexts: The Hungarian Educational Longitudinal Program. Keynote address. Children and Youth in Changing Societies. 3rd International Conference. Thessaloniki, Greece, December 2-4, 2010. 170.

Csapó, B. (2002). Az új tudás képződésének eszközei: az induktív gondolkodás [Tools of generating new knowledge: inductive reasoning]. In B. Csapó (Ed.), *Az iskolai tudás*. [School knowledge]. 2nd edition. (pp. 261–290). Budapest: Osiris Kiadó.

Csapó, B. (1997). The development of inductive reasoning: cross-sectional assessments in an educational context. *International Journal of Behavioral Development, 20*(4), 609–626. https://doi.org/10.1080/016502597385081

Csapó, B., Ainley, J., Bennett, R., Latour, T., & Law, N. (2012). Technological issues of computer-based assessment of 21st century skills. In B. McGaw, P. Griffin, & E. Care (Eds.), *Assessment and teaching of 21 century skills* (143–230). New York, NY: Springer. https://doi.org/10.1007/978-94-007-2324-5_4

Csapó, B. & Csépe, V. (2012). *Framework for diagnostic assessment of reading*. Budapest: Nemzeti Tankönyvkiadó.

Csapó, B. & Csíkos, C. (2011). Diagnostic assessment frameworks for mathematics: Theoretical background and practical issues. In B. Csapó & M. Szendrei (Eds.), *Framework for diagnostic assessment of mathematics* (pp. 137–162). Budapest: Nemzeti Tankönyvkiadó.

Csapó, B., Csíkos, C., & Molnár, G. (Eds.). (2015). *A matematikai tudás online diagnosztikus értékelésének tartalmi keretei* [Framework for online diagnostic assessment of mathematics]. Budapest: Nemzeti Tankönyvkiadó.

Csapó, B., Korom, E., & Molnár, G. (Eds.). (2015). *A természettudományi tudás online diagnosztikus értékelésének tartalmi keretei* [Framework for online diagnostic assessment of science]. Budapest: Nemzeti Tankönyvkiadó.

Csapó, B., Lőrincz, A., & Molnár, G. (2012). Innovative assessment technologies in educational games designed for young students. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in Game-Based Learning* (pp. 235–254), Springer, New York. https://doi.org/10.1007/978-1-4614-3546-4_13

Csapó, B. & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology, 10*(1522). Retrived from https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01522/full

Csapó, B., Molnár, G., & Kinyó, L. (2008). Analysis of the selectiveness of the Hungarian educational system in international context. The 3rd IEA International Research Conference, Taipei, 16–20 September.

Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology, 106*(3), 639–650. https://doi.org/10.1037/a0035756

Csapó, B, Molnár, G., & R. Tóth, K. (2009). Comparing paper-and-pencil and online assessment of reasoning skills: A pilot study for introducing TAO in large-scale assessment in Hungary. In F. Scheuermann, & J. Björnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 113–118). Luxemburg: Office for Official Publications of the European Communities.

Csapó, B. & Nikolov, M. (2009). The cognitive contribution to the development of proficiency in a foreign language. *Learning and Individual Differences, 19*(2), 209–218. https://doi.org/10.1016/j.lindif.2009.01.002

Csapó, B., Steklács, J., & Molnár, G. (Eds.). (2015). *Az olvasás-szövegértés online diagnosztikus értékelésének tartalmi keretei* [Framework for online diagnostic assessment of reading]. Budapest: Nemzeti Tankönyvkiadó.

Csapó, B. & Szabó, G. (2012). *Framework for diagnostic assessment of science*. Budapest: Nemzeti Tankönyvkiadó.

Csapó, B. & Szendrei, M. (2011). *Framework for diagnostic assessment of mathematics*. Budapest: Nemzeti Tankönyvkiadó.

de Konig, E. (2000). *Inductive reasoning in primary education: Measurement, teaching, transfer*. Zeist (Netherlands): Kerckebosch.

De Koning, E., & Hamers, J. H. M. (1999). Teaching inductive reasoning: Theoretical background and educational implications (pp. 157–188). In J. H. M Hamers, J. E. H. Van Luit, & B. Csapó (Eds.), *Teaching and learning thinking skills*. Lisse: Swets & Zeitlinger.

de Koning, E., Hamers, Jo, H. M., Sijtsma, K., & Vermeer, A. (2002). Teaching inductive reasoning in primary education. *Developmental Review*, 22(2), 211–241. https://doi.org/10.1006/drev.2002.0548

Dienes, Z. (1973). (Building up mathematics) *Építsük fel a matematikát*. Budapest: Gondolat Kiadó.

Dienes, Z. P. (1963). *An experimental study of mathematics learning*. London: Hutchinson.

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning, and Assessment, 5*(1). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1640/1489

Dormann, C. & Biddle, R. (2006). Humour in game-based learning. *Learning, Media & Technology, 31*(4), 14.

Downs, R. & de Souza, A. (2005). *Learning to think spatially: GIS as a support system in the K-12 Curriculum*. Washington D.C.: National Research Council and National Academies Press. DOI:10.17226/11019

Ebner, M. & Holzinger, A. (2007). Successful implementation of user-centred game-based learning in higher education: an example from civil engineering. *Computers & Education, 49*(3), 873–890.

Egan, G. E. & Greeno, J. G. (1974). Theory of rule induction: Knowledge acquired in concept learning, serial pattern learning and problem solving. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 43–104). Potomac, MD: Lawrence Erlbaum Associates Inc.

Egenfeldt-Nielsen, S. (2007). Making sweet music: The educational use of computer games. In J. Linderoth (Ed.), *Datorspelandets Dynamik. Studentlitteratur*. Retrieved from http://egenfeldt.eu/papers/sweet_music.pdf

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron, & R. J. Sternberg (Eds.), *Teaching thinking skills* (pp. 9–26). New York: W.H. Freeman.

Farcot, M. & Latour, T. (2009). Transitioning to computer-based assessments: A question of costs. In F. Scheuermann & J. Bjornsson (Eds.), *The transition to Computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 108–116). Brussels: European Communities.

Farcot, M. & Latour, T. (2008). An open source and large - scale computer-based assessment platform: A real winner. In F. Scheuermann & A. G. Pereira (Eds.), *Towards a research agenda on computer - based assessment: Challenges and needs for European educational measurement* (pp. 64–67). Ispra: European Commission Joint Research Centre.

Feuerstein, R., Rand, Y., Hoffman, M. B., & Miller, R. (1980). *Instrumental enrichment: An intervention programme for cognitive modifiability*. Baltimore, MD: University Park Press.

Fishbein, B., Martin, M. O., Mullis, I. V. S., & Foy, P. (2018). The TIMSS 2019 Item Equivalence Study: examining mode effects for computer-based assessment and implications for measuring trends. *Large-Scale Assessments in Education*, *6*(1), doi: 10.1186/s40536-018-0064-z

Frey, A. (2007). Adaptives Testen. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Testkonstruktion* (pp. 261–278). Berlin, Heidelberg: Springer.

Funke, J. & Frensch, P. A. (2007). Complex problem solving: The European perspective – 10 years after. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems* (pp. 25–47). New York: Erlbaum.

Gallagher, A., Bridgeman, B., & Cahalan, C. (2000). *The effect of computer-based tests on racial/ethnic, gender, and language groups* (GRE Board Professional Report No. 96-21P). Princeton, NJ: Education Testing Service.

Garris, R., Ahlers, R., & Driskell, J. (2002). Games, motivation, and learning: a research and practice model. *Simulation and Gaming, 33*(4), 441–467.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave Macmillan.

Gelman, S. A. & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58*, 1532–1541.

Gibney, J. (2014). Provoking mathematical thinking: experiences of doing realistic mathematics tasks with adult numeracy teachers. *Adults Learning Mathematics*, *9*(2), 97–115.

Gick, M. L. & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive psychology, 15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Gilhooly, K. J. (1982). *Thinking: Directed, undirected and creative.* London: Academic Press.

Goldman, S. R. & Pellegrino, J. W. (1984). Deductions about induction: Analyses of developmental and individual differences. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol. 2* (pp. 149–197). Hillsdale, NJ: Erlbaum.

Goldman, S. & Pellegrino, J. (1982). Development and individual differences in verbal analogical reasoning. *Child Development, 53*(1), 550–559. https://doi.org/10.2307/1128998

Goswami, U. (1991). Analogical reasoning: what develops? A review of research and theory. *Child Development*, *62*(1), 1–22. https://doi.org/10.2307/1130701

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement*, 36, 189–213. https://doi.org/10.1177/0146621612439620

Griffin, P., McGaw, B., & Care, E. (Eds.). (2012). *Assessment and teaching of 21st century skills*. New York: Springer.

Hamers, J. H. M., De Koning, E., & Sijtsma, K. (2000). Inductive reasoning in the third grade: Intervention promises and constraints. *Contemporary Educational Psychology*, *23*, 132–148. https://doi.org/10.1006/ceps.1998.0966

Hamers, J. H. M., De Koning, E., & Sijtsma, K. (1998). Inductive reasoning in third grade: Intervention promises and constraints. *Contemporary Educational Psychology*, *23*(2), 132-148.

Hattie, J. (2009). *Visible learning: A synthesis of 800 meta-analyses relating to achievement.* New York, NY: Routledge.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.

Hattie, J., & Anderman, E. M. (Eds.). (2013). *International guide to student achievement*. Routledge.

Hayes, B., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(2), 278–292. https://doi.org/10.1002/wcs.44

Heit, E. (2001). What is induction and why study it? In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental, and computational approaches*. Cambridge: Cambridge University Press.

Herlina, E. (2015). Advanced mathematical thinking and the way to enhance IT. *Journal of Education and Practice, 6*(5), 79–89.

Hotulainen, R., Mononen, R., & Aunio, P. (2016). Thinking skills intervention for low-achieving first graders. *European Journal of Special Needs Education, 31*(3), 360–375.

Huang, W. Y. & Wong, S. H. (2014). Cross-cultural validation. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 1369–1371). Dordrecht: Springer.

Hudson, B., Henderson, S., & Hudson, A. (2015). Developing mathematical thinking in the primary classroom: Liberating students and teachers as learners of mathematics. *Journal of Curriculum Studies, 47*(3), 374–398. https://doi.org/10.1080/00220272.2014.979233

Iacovides, I., Aczel, J., Scanlon, E., & Woods, W. (2012). Investigating the relationships between informal learning and player involvement in digital games. *Learning, Media and Technology, 37*(3), 321–327.

Ifenthaler, D., Eseryel, D., & Ge, X. (2012). Assessment for game-based learning. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning. Foundations, innovations, and perspectives* (pp. 3–10). New York, NY: Springer.

Jensen, I. L. (1973). Acts: An Inductive Study. Moody Publishers.

Jodoin, M., Zenisky, A., & Hambleton, R. K. (2006). Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes. *Applied Measurement in Education, 19*(3), 203–220.

Johnson- Laird, P. N. (1988). A taxonomy of thinking. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought* (pp. 429–457). Cambridge: Cambridge University Press.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice*, *22*(1), 36–46.

Jurecka, A. & Hartig, J. (2007). Computer- und Netzbasiertes Assessment. In J. Hartig & E. Klieme (Eds.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (pp. 37–48). Berlin, Bonn: Bundesministerium für Bildung und Forschung.

Kambeyo, L. (2017). The possibilities of assessing students' scientific inquiry skills using an online instrument: A small-scale study in the Omusati Region, Namibia. *European Journal of Educational Sciences, 4*(2), 1–21.

Kikis, K. (2010). Reflections on paper-and-pencil tests to eAssessments: Narrow and broadband paths to 21st Century challenges. In F. Scheuermann & J. Bjornsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large- scale testing* (pp. 99–103). Brussels: European Communities.

Kim, Y. J. & Ifenthaler, D. (2019). Game-based assessment: The past ten years and moving forward. *Advances in Game-Based Learning Game-Based Assessment Revisited*, 3–11. https://doi.org/10.1007/978-3-030-15569-8_1

Kingston, N. (2008). Comparability of computer and paper administered multiple choice tests for K-12 populations: a synthesis. *Applied Measurement in Education, 22*(1), 22–37. https://doi.org/10.1080/08957340802558326

Klauer, K. (1996). Begünstigt induktives Denken das Lösen komplexer Probleme? Experimentellen Studien zu Leutners sahel-Problem [Does inductive thinking have a positive effect on complex problem solving?]. *Zeitschrift für Experimentelle Psychologie, 43*(1), 361–366 (Dutch version).

Klauer, K. (1993). *Denktraining für Jugendliche* [Cognitive Training for Youth]. Göttingen: Hogrefe (Dutch version).

Klauer, K. (1991). *Denktraining für Kinder II* [Cognitive Training for Children II]. Göttningen: Hogrefe (Dutch version).

Klauer, K. J. (1990). Paradigmatic teaching of inductive thinking. In H. Mandl, E. De Corte, S. N. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction. European Research in an International context. Analysis of complex skills and complex knowledge domains* (pp. 23–45). Oxford: Pergamon Press.

Klauer, K. (1989). *Denktraining für Kinder* [Cognitive training for children 1]. Göttingen: Hogrefe (German version).

Klauer, K. J. & Phye, G. D. (1994). *Cognitive training for children. A developmental program of inductive reasoning and problem solving.* Seattle, WA: Hogrefe & Huber.

Klauer, K. & Phye, G. (2008). Inductive reasoning: a training approach. *Review of Educational Research*, *78*(1), 85–123. Retrived from https://doi.org/10.3102/0034654307313402

Klauer, K., Willmes, K., & Phye, G. (2002). Inducing inductive reasoning: does it transfer to fluid intelligence? *Contemporary Educational Psychology*, *27*(1), 1–25. https://doi.org/10.1006/ceps.2001.1079

Klawe, M. M. (1998). *When does the use of computer games and other interactive multimedia software help students learn mathematics?* Unpublished manuscript.

Kline, R. B. (2016). *Principles and practice of structural equation modelling, 4th Ed.* New York, NY: The Guilford Press.

Kozma, R. (2009). Assessing and teaching 21st century skills: A call to action. In F. Scheuermann, & J. Bjornsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large scale assessment* (pp. 13–23). Brussels: European Communities.

Levin, J. (1981). Estimation techniques for arithmetic: Everyday math and mathematics instruction. *Educational Studies in Mathematics*, *12*, 421–434.

Lipman, M. (1985). Thinking skills fostered by Philosophy for Children. In J. W. Segal, S. F. Chipman, dan R. Glaser (Eds.), *Thinking and learning skills* (pp. 83–108). Hillsdale, NJ: Lawrence Erlbaum Associates.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Struct. Equ. Modeling, 9*, 151–173. doi: 10.1207/S15328007SEM0902_1

Magyar, A. & Molnár, G. (2015). A szóolvasási készség online mérésére kidolgozott adaptív és lineáris tesztrendszerek összehasonlító hatékonyságvizsgálata [Adaptive measurement of reading competence: An effectiveness study]. *Magyar Pedagógia, 115*(4), 403–428.

Markman, E. M. (1989). *Categorization and naming in children. Problems of Induction*. Cambridge, MA: MIT Press.

Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in science*. Retrieved from Boston College, TIMSS & PIRLS

International Study Center website: http://timss andpi rls.bc.edu/timss 2015/inter natio nal-resul ts/.

Martin, R. (2010). Utilising the potential of Computer delivered surveys in assessing scientific Literacy. In F. Scheuermann & J. Bjornsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 172–177). Brussels: European Communities.

Martin, R. & Binkley, M. (2009). Gender differences in cognitive tests: A consequence of gender-dependent preferences for specific information presentation formats? In F. Scheuermann & J. Bjórnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 75–82). Luxembourg: Office for Official Publications of the European Communities.

Mayer, R. E. (1992). *Thinking, problem solving, cognition* (2nd ed.). New York: Freeman.

Mayrath, M. C., Clarke-Midura, J., & Robinson, D. H. (2012). *Introduction to Technology-based assessments for 21st century skills.* In M. C. Mayrath, J. Clarke-Midura, D. H. Robinson, & G. Schraw (Eds.), *Current perspectives on cognition, learning, and instruction. Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (p. 1–11). Information Age Publishing, Inc.

McGuinness, C. (2007). Building thinking skills in thinking classrooms: ACTS in Northern Ireland. 109–114. Paper presented at 13th International Conference on Thinking, Norrkoping, Sweden.

McGuinnes, C., Sheey, N., Curry, C., & Eakin, A. (2003). *ACTs II Sustainable thinking in classrooms: A methodology for enhancing thinking across the curriculum*. School of Psychology, Queen's University, Belfast, Northern Ireland.

Meijer, R. (2010). Transition to computer-based assessment: Motivations and considerations. In F. Scheuermann & J. Bjornsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 104–107). Brussels: European Communities.

Meluso, A., Zheng, M., Spires, H., & Lester, J. (2012). Enhancing 5th graders' science content knowledge and self-efficacy through game-based learning. *Computers & Education*, *59*(2), 497–504.

Mitchell, T., Russel, T., Broomhead, P., & Aldridge, N. (2002). Towards robust computerized marking of free-text responses. In M. Danson (Ed.), *Proceedings of the sixth International computer assisted assessment conference*. Loughborouh: Loughboroug University. Retrieved from https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/1884/1/Mitchell_t1.pdf

Moe, E. (2010). Introducing large-scale computerized assessment – Lessons learned and future challenges. In F. Scheuermann & J. Bjórnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 51–56). Luxembourg: Office for Official Publications of the European Communities.

MoEHE. (2017). Basic characteristics of the Palestinian general educational system. Retrieved from http://www.moehe.gov.ps/en/About-the-Ministry/Education-System

Molnár, G. (2011). Playful fostering of 6- to 8-year-old students' inductive reasoning. *Thinking Skills and Creativity*, *6*(2), 91–99. doi: 10.1016/j.tsc.2011.05.002

Molnár, G. (2006). Az induktív gondolkodás fejlesztése kisiskolás korban [Fostering inductive thinking in elementary school pupils]. *Magyar Pedagógia, 1*, 63–80.

Molnár, G. & Csapó, B. (2013). Az eDia online diagnosztikus mérési rendszer [The eDia online diagnostic measurement system] XI. Pedagógiai Értékelési Konferencia. Szeged, 2012, 82, pp. 11–13.

Molnár, G., & Csapó, B. (2011). Az 1–11 évfolyamot átfogó induktív gondolkodás kompetenciaskála készítése a valószínűségi tesztelmélet alkalmazásával [Using item response theory to design an inductive thinking competence scale for the first to eleventh grades]. *Magyar Pedagógia, 111*(2), 127–140.

Molnár, G. & Csapó, B. (2019a). How to make learning visible through technology: The eDia-Online Diagnostic Assessment System. In *Proceedings of the 11th International Conference on Computer Supported Education* (CSEDU 2019) Volume 2, pp. 122–131.

Molnár, G. & Csapó, B. (2019b). Technology-based diagnostic assessments for identifying early mathematical learning difficulties. *International Handbook of Mathematical Learning Difficulties*, 683–707. doi: 10.1007/978-3-319-97148-3_40

Molnár, G. & Csapó, B. (2019c). Making the psychological dimension of learning visible: Using technology-based assessment to monitor students' cognitive development. *Frontiers in Psychology*, *10*. doi: 10.3389/fpsyg.2019.01368

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, 9(1), 35–45. https://doi.org/10.1016/j.tsc.2013.03.002

Molnár, G., Greiff, S., Wüstenberg, S., & Fischer, A. (2017). Empirical study of computer-based assessment of domain-general dynamic problem solving skills. In B. Csapó, J. Funke, & A. Schleicher (Eds.), *The nature of problem solving* (pp. 123–143). Paris: OECD.

Molnár, G. & Lőrincz, A. (2012). Innovative assessment technologies: Comparing 'face-to face' and game-based development of thinking skills in classroom settings. 2012 International Conference on Management and Education Innovation, 37, pp. 150–154.

Molnár, G. & Pásztor, A. (2015a). Feasibility of computer-based assessment at the initial stage of formal schooling: The developmental level of keyboarding and mouse skills in Year One. Conference on Educational Assessment, Hungary, April 23–25, 2015, p. 118.

Molnár, G. & Pásztor, A. (2015b). A számítógép alapú mérések megvalósíthatósága kisiskolás diákok körében: első évfolyamos diákok egér- és billentyűzet-használati képességeinek fejlettségi szintje [The feasibility of computer-based measurements among lower primary school students: The developmental level of Year 1 students' keyboarding and mouse skills]. *Magyar Pedagógia, 115*(3), 237-252.

Molnár, G., Pásztor, A., & Csapó, B. (2019): The eLea online training platform. In Molnár, E. K. & Dancs, K. (Eds.), XVII. Pedagogical Assessment Conference - 17th Conference on Educational Assessment. Program and abstracts. University of Szeged, Szeged. 64.

Mousa, M. & Molnár, G. (2019a). The feasibility of computer-based testing in Palestine among lower primary school students: Assessing mouse skills and inductive reasoning. *Journal of Studies in Education*, *9*(2), 1–16. doi:10.5296/jse.v9i2.14517

Mousa, M. & Molnár, G. (2019b). Applying computer-based testing in Palestine: Assessing fourth and fifth graders inductive reasoning. *Journal of Studies in Education*, *9*(3), 1–13. doi:10.5296/jse.v9i3.14799

Mousa, M. & Molnár, G. (2018). Introducing computer-based assessment among 2nd to 4th grade students in Palestine. 16th Conference on Educational Assessment, 16,71.

Mullis, I. V. S. & Martin, M. O. (Eds.). (2019). *PIRLS 2021 Assessment Frameworks*. Boston College Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timss andpi rls.bc.edu/timss 2015/inter natio nal-resul ts/.

Mullis, I. V. S. & Martin, M. O. (Eds.). (2013). *TIMSS 2015 assessment frameworks*. Boston College Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Muthén, L. K. & Muthén, B. O. (2010). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.

Nagy, J. (2008). Az alsó tagozatos oktatás megújítása, In K. Fazekas, J. Köllő J., & J. Varga (Eds.), *Zöld Könyv a magyar közoktatás megújításáért*. Ecostat, Budapest, 53–70 (Hungarian version).

Nikolov, M. & Csapó, B. (2018). The relationships between 8th graders' L1 and L2 readings skills, inductive reasoning and socio-economic status in early English and German as a foreign language programs. *System*, *73*, 184–206.

Nisbet, J. (1993). The thinking curriculum. *Educational psychology*, *13*(3–4), 281–290.

OECD. (2016). *PISA 2015 results* (Volume I). *Excellence and equity in education*. Paris.

OECD. (2013). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. Paris: OECD. doi: 10.1787/9789264190511-en

Onal, H., Inan, M., & Bozkurt, S. (2017). A research on mathematical thinking skills: Mathematical thinking skills of athletes in individual and team sports. *Journal of Education and Training Studies*, *5*(9), 133. doi: 10.11114/jets.v5i9.2428

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological review, 97*(2), 185. https://doi.org/10.1037/0033-295X.97.2.185

Pacetti, E. (2008). Improving the quality of education in Palestine through e-Learning and ICT: The bottom-up approach for a sustainable pedagogy. In *Proceedings of the Conference Knowledge Construction in E-learning Context: CSCL, ODL, ICT and SNA in education.* Cesena, Italy.

Pachler, N., Daly, C., Mor, Y., & Mellar, H. (2010). Formative e-assessment: Practitioner cases. *Computers and Education, 54*(3), 715–721.

Pásztor, A. (2014). Playful fostering of inductive reasoning through mathematical content in computer-based environment. 12th Conference on Educational Assessment, Szeged, May 1-3. 2014. p93.

Pásztor, A. (2015). Computer based assessment and development of inductive reasoning strategies. Szeged workshop on educational evaluation, Szeged. Abstract Retrieved on March 2017, from: http://www.edu.uszeged.hu/swee/eng/2015/SWEE_7_programfuzet_absztraktokkal.pdf

Pásztor, A. (2016). Technology-based assessment and development of inductive reasoning. Ph.D. thesis. Doctoral School of Education, University of Szeged. Retrieved from http://www.edu.u-szeged.hu/phd/people/apasztor/public_en_all.html

Pásztor, A., Hotulainen, R., Kupiainen, S., Molnár, G., & Csapó, B. (2018). Comparing Finnish and Hungarian fourth grade students' inductive reasoning skills. EARLI SIG 1 Conference, Helsinki.

Pásztor, A., Molnár, G., Korom, E., B. Németh, M., & Csapó, B. (2017). Online assessment of inductive reasoning and its predictive power on inquiry skills in science. 17th biennial conference of the European Association for Research on Learning and Instruction (EARLI). p509. Tampere, Finland, August 29-September 2, 2017.

Molnár, G. & Pásztor, A. (2012): Game-based development of thinking skills. Paper presented at the 4th Szeged Workshop on Educational Evaluation (SWEE), Szeged, 24-25. April 2012.

Pásztor, A. & Molnár, G. (2013). Validity and effectiveness of computer-based versus face-to-face testing in first grade. Paper presented at the 15th European Conference for the Research on Learning and Instruction. Munich, Germany, 27-31. August 2013. p64.

Pásztor, A., Molnár, G., & Csapó, B. (2015). Technology-based assessment of creativity in educational context: The case of divergent thinking and its relation to mathematical achievement. *Thinking skills and Creativity*, *18*, 32–42.

PCBS. (2013). The woman and the man in Palestine, issues and statistics. *The Palestinian central bureau for statistics.* Ramallah, Palestine (Arabic version). Retrieved from: http://www.pcbs.gov.ps/Portals/_PCBS/Downloads/book2015.pdf

Peak, P. (2005). *Recent trends in comparability studies*. Pearson educational measurement. Retrieved from http://www.pearsonassessments.com/NR/rdonlyres/5FC04F5A-E79D-45FE848407AACAE2DA75/0/TrendsCompStudies_rr0505.pdf

Pearson (2012). From paper and pencil to computer-based testing (CBT). Retrieved from http://www.pearsonvue.co.uk/india/Documents/PP_to_CBT.pdf.

Pellegrino, J. W. & Glaser, R. (1982). Analyzing aptitudes for learning: Inductive reasoning. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 269–345). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perret, P. (2015). Children's inductive reasoning: Developmental and educational perspectives. *Journal of Cognitive Education and Psychology*, *14*(3), 389–408. doi: 10.1891/1945-8959.14.3.389.

Phye, G. D. (1989). Schemata training and transfer of an intellectual skill. *Journal of Educational Psychology*, *81*, 347–352.

Pivec, M. (2007). Editorial: play and learn: potentials of game-based learning. *British Journal of Educational Technology, 38*(3), 387–393.

Plichart, P., Jadoul, R., Vandenabeele, L., & Latour, T. (2004). TAO, A collective distributed computer-based assessment framework built on semantic web standards. In *Proceedings of the international conference on Advances in intelligent Systems – Theory and application AISTA2004, in cooperation with IEEE computer society*. 2004 Nov 15–18; Luxembourg. Luxembourg.

PMEHE (2018). *Educational Statistics 2016/2017*. Palestine Ministry of Education and Higher Education, Ramallah. Retrieved from http://www.mohe.pna.ps/services/statistics

Prensky, M. (2007). *Digital game-based learning*. Saint Paul, MN: Paragon House.

Prensky, M. (2001). *Digital game-based learning*. New York, NY: McGraw-Hill.

Price, P., Tepperman, J., Iseli, M., Duong, T., Black, M., Wang, S., & Alwan, A. (2009). Assessment of emerging reading skills in young native speakers and language learners. *Speech Communication, 51*(1), 968–984.

Rajendran, N. S. (2010). *Teaching and acquiring higher order thinking skills: Theory and practice*. Tanjong Malim, Perak: Penerbit Universiti Pendidikan Sultan Idris.

Ramaley, J. A. (2007). Aims of mathematics education. *Assessing Mathematical Proficiency*, 17–22. doi: 10.1017/cbo9780511755378.004

Rausch, A. & Pásztor, A. (2017). Exploring the possibilities of online assessment of early numeracy in kindergarten. In *Proceedings of the 41st Conference of the International Group for the Psychology of Mathematics Education,* Vol. 4, pp. 89–96. Singapore: PM.

Resing, W. C. M. (1993). Measuring inductive reasoning skills: The construction of a learning potential test. In J. H. M. Hammers, K. Sijstma, & A. J. J. Ruijssenaars (Eds.), *Learning potential assessment: Theoretical, methodological and practical issues* (pp. 219–242). Amsterdam: Swets and Zeitlinger.

Resnick, L. (1987). *Education and learning to think.* Washington, DC: National Academy Press.

Resnick, L. B. & Klopfer, L. E. (1989). *Toward the thinking curriculum: Current cognitive research. 1989 ASCD Yearbook.* Alexandria, VA: Association for Supervision and Curriculum Development.

Riel, M. (1994). Educational change in a technology-rich environment. *Journal of Research on Computing in Education, 26*(4), 452–474.

Ridgway, J. & McCusker, S. (2003). Using computers to assess new educational goals. *Assessment in Education, 10*(3), 309–328.

Ripley, M. (2010). Transformational computer-based testing. In F. Scheuermann & J. Bjórnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 92–98). Luxembourg: Office for Official Publications of the European Communities.

Ritchhart, R., Church, M., & Morrison, K. (2011). *Making thinking visible: How to promote engagement, understanding, and independence for all learners*. San Francisco: Jossey-Bass.

Robinson, I. (1987). *Program to incorporate high-order thinking skills into teaching and learning for grades K-3*. Fort Lauderdale, FL: Nova University. ED 284 689).

Ropo, E. (1987). Skills for learning: a review of studies on inductive reasoning. *Scandinavian Journal of Educational Research*, *31*(1), 31–39.

Rosas, R. e. a. (2003). Beyond Nintendo: A design and assessment of educational video games for first and second grade students. *Computers & Education*, *40*, 71–94.

Sandberg, E. & McCullough, M. (2010). The development of reasoning skills. In Sandberg, E. & Spritz, B. (Eds.), *A clinician's guide to normal cognitive development in childhood* (179–189). New York: Routledge. https://doi.org/10.4324/9780203843697

Saragih, S. & Napitupulu, E. (2015). Developing student-centered learning model to improve high order mathematical thinking ability. *International Education Studies*, *8*(6), 104–112. https://doi.org/10.5539/ies.v8n6p104

Scheuermann, F. & Björnsson, J. (2009). *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. Luxemburg: Office for Official Publications of the European Communities.

Scheuermann, F. & Pereira, A. G. (2008). *Towards a research agenda on computer-based assessment. Challenges and needs for European educational measurement*. Luxembourg: Office for Official Publications of the European Communities.

Sedighian, K. & Sedighian, A. S. (1996). *Can educational computer games help educators learn about the psychology of learning mathematics in children?* Paper presented at the 18th Annual Meeting of the International Group for the Psychology of Mathematics Education, Florida.

SETDA. (2008). *National trends report 2008. Enhancing education through technology*. Retrieved from http://www.setda.org/wp-content/uploads/2013/12/National-Trends-Report-2008.pdf.

Shaffer, D. W. (2006). *How computer games help children learn?* New York, NY: Palgrave Macmillan.

Shayer, M. (1999). Cognitive acceleration through science education II: Its effect and scope. *Int. J. Sci. Educ. 21*, 883–902.doi: 10.1080/095006999290345

Shayer, M., & Adey, P. (Eds.). (2002). *Learning intelligence: Cognitive acceleration across the curriculum from 5 to 15 years.* Milton Keynes: Open University Press.

Shayer, M., & Adey, P. (1981). *Towards a science of science teaching: Cognitive development and curriculum demand*. London: Heinemann Educational Books.

Shayer, M. & Adhami, M. (2010). Realizing the cognitive potential of children 5–7 with a mathematics focus: Post-test and long-term effects of a 2-year intervention. *British Journal of Educational Psychology, 80*(3), 363–379.

Shayer, M. & Adhami, M. (2007). Fostering cognitive development through the context of mathematics: results of the CAME project. *Educ. Stud. Math, 64*, 265–291.doi: 10.1007/s10649-0069037-1

Shihab, H. (2014). *It aims to improve the educational techniques in Palestine. "My book Net" the educational digital book for student.* (Arabic version) Retrieved from: https://www.alwatanvoice.com/arabic/content/print/617249.html

Shraim, K. (2018). Palestine (West Bank and Gaza Strip). In A. Weber & S. Hamlaoui (Eds), *E-Learning in the Middle East and North Africa (MENA) Region* (pp 309–332). Springer, Cham. Retrieved from https://doi.org/10.1007/978-3-319-68999-9_14

Shye, S. (1988). Inductive and deductive reasoning: a structural reanalysis of ability tests. *Journal of Applied Psychology, 73*(2), 308–311.

Sim, G. & Horton, M. (2005). Performance and attitude of children in Computer based versus paper-based testing. In P. Kommers & G. Richards (Eds.), *Proceedings of world conference on educational multimedia, hypermedia and Telecommunications 2005* (pp. 3610–3614). Chesapeake, VA: AACE.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 231–280. https://doi.org/10.1006/cogp.1993.1006

Sloman, S. & Lagnado, D. (2005). *The problem of induction. The Cambridge handbook of thinking and reasoning*. Cambridge University Press, 95–116.

Söderqvist, S., Nutley, S. B., Ottersen, J., Grill, K. M., & Klingberg, T. (2012). Computerized training of non-verbal reasoning and working memory in children with intellectual disability. *Frontiers in Human Neuroscience,* 6:271. doi:10.3389/fnhum.2012.00271

Sternberg, R. J. (1986). Toward a unified theory of human reasoning. *Intelligence*, *10*, 281–314.

Sternberg, R. J. & Gardner, M. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General, 112*, 80–116.

Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods Psychol. Res. Online, 2*, 21–33.

Strain-Seymour, E., Way, W. D., & Dolan, R. P. (2009). *Strategies and processes for developing innovative items in large-scale assessments*. Iowa City, IA: Pearson Education.

Swartz, R. (2001). Infusing critical and creative thinking into content instruction. In A. L. Costa (Ed.), *Developing minds: A resource book for teaching thinking* (pp.266-274) 3rd Ed. Alexandria, VA: Association for Supervision and Curriculum Development.

Thurlow, M., Lazarus, S. S., Albus, D., & Hodgson, J. (2010). *Computer-based testing: Practices and considerations (Synthesis Report 78)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Tomic, W. (1995). Training in inductive reasoning and problem solving. *Contemporary Educational Psychology*, *20*(1), 483–490. https://doi.org/10.1006/ceps.1995.1036

Tomic, W., & Kingma, J. (Eds.). (1998). *Conceptual issues in research in intelligence*. JAI Press.

Tomic, W. & Klauer, K. J. (1996). On the effects of training inductive reasoning: How far does it transfer and how long do the effects persist? *European Journal of Psychology of Education*, *3*, 283–299.

UNRWA (2017). Who we are. The United Nations relief and works agency for Palestine refugees. Retrieved from https://www.unrwa.org/who-we-are

Vainikainen, M.-P., Hautamaki, J., Hotulainen, R., & Kupiainen, S. (2015). General and specific thinking skills and schooling: Preparing the mind to new learning. *Thinking Skills and Creativity, 18*, 53–64. https://doi.org/10.1016/j.tsc.2015.04.006

Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education: Research*, *2*(1), 319–330.

Van de Vijver, F. J. R. (1991). *Inductive thinking across cultures: An empirical investigation.* Tilburg (Netherlands): Catholic Univ. of Brabant.

van Lent, G. (2010). Risks and benefits of CBT versus PBT in high-stakes testing. In F. Scheuermann & J. Bjornsson (Eds.), *The transition to computer-based* *assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 83–91). Brussels: European Communities.

Vo, D. V., & Csapó, B. (2020). Development of inductive reasoning in students across school grade levels. *Thinking Skills and Creativity, 37*, 100699. doi:10.1016/j.tsc.2020.100699

Voogt, J., Erstad, O., Dede, C., & Mishra, P. (2013). Challenges to learning and schooling in the digital networked world of the 21st century. *Journal of Computer Assisted Learning*, *29*(5), 403–413.

Way, W., Davis, L., & Fitzpatrick, S. (2006). Score comparability of online and paper administrations of Texas assessment of knowledge and skills. Paper presented at the annual meeting of the National Council on Measurement in Education. San Francisco, CA.

Weaver, A. J. & Raptis, H. (2001). Gender differences in introductory atmospheric and oceanic science exams: multiple-choice versus constructed response questions. *Journal of Science Education and Technology*, *10*(2), 115–127.

Wegerif, R. (2006). A dialogic understanding of the relationship between CSCL and teaching thinking skills. *International Journal of Computer-Supported Collaborative Learning*, *1*(1), 143–157.

Wegerif, R., Li, L., & Kaufman, J. C. (2017). *The Routledge international handbook of research on teaching thinking.* Oxford: Routledge.

Werts, M., Caldwell, N. & Wolery, M. (2003). Instructive feedback: effects of a presentation variable. *The Journal of Special Education*, *37*(2), 124–133.

Williamson, D. M., Mislevy, R. J., & Bejar, I. I. (Eds.). (2006). *Automated scoring of complex tasks in computer-based testing*. Mahwah: Erlbaum.

World TVET Database Palestine (2015). UNESCO-UNEVOC International Centre for Technical and Vocational Education and Training. Retrieved from https://unevoc.unesco.org/wtdb/worldtvetdatabase_pse_en.pdf

Wu, H. & Molnár G. (2018a). Computer-based assessment of Chinese students' component skills of problem solving: A pilot study. *International Journal of Information and Education Technology*, *8*(5), 381–356. https://doi.org/10.18178/ijiet.2018.8.5.1067

Wu, H. & Molnár, G. (2018b). Interactive problem solving: Assessment and relations to combinatorial and inductive reasoning. *Journal of Psychological and Educational Research*, *26*(1), 90–105.

Yang, Y. & Chang, C. (2013). Empowering students through digital game authorship: enhancing concentration, critical thinking, and academic achievement. *Computers & Education, 68*(1), 334–344.

Yang, Y. (2012). Building virtual cities, inspiring intelligent citizens: digital games for developing students' problem solving and learning motivation. *Computers & Education, 59*(1), 365–377.

Yelland, N. (2005). The future is now: A review of the literature on the use of computers in early childhood education (1994–2004). *AACE Journal*, *13*(3), 201–232.

Yesildere, S. & Turnuklu, E. B. (2007). Examination of students' mathematical thinking and reasoning processes. *Journal of Faculty of Educational Sciences, 40*(1), 181–213. https://doi.org/10.1501/Egifak_0000000152

Yuda, M. (2011). Effectiveness of digital educational materials for developing spatial thinking of elementary school students. *Procedia - Social and Behavioral Sciences*, *21*(1), 116–119.

10 APPENDICES

10.1 Appendix A: Permission letters from the Ministry of Education and Higher Education - Directorate of Education / Bethlehem

بسم الله الرحمن الرحيم

State of Palestine
Ministry of Education & Higher Education
Directorate of Education\Bethlehem

دولة فلسطين
وزارة التربية والتعليم العالي
مديرية التربية والتعليم/بيت لحم

الرقم: 3 / 1 / 2
التاريخ: 2017.01.02م
الموافق: 4 ربيع الثاني 1438 هـ

مديري ومديرات المدارس الحكومية والخاصة المحترمين
تحية طيبة وبعد،،،

الموضوع: تسهيل مهمة

لا مانع من تسهيل مهمة الطالب **مجاهد علي** موسى، والسماح له بتطبيق دراسته بعنوان: "Enhancing first graders thinking skills in Palestine"، خلال الفصل الدراسي الأول والثاني للعام الدراسي 2017/2016م، على أن لا يؤثر ذلك على سير العملية التعليمية، علما بأن المعلومات لن تستخدم إلا لأغراض البحث العلمي.

مع الاحترام

أ. سامي كامل مروّة
/ مدير التربية والتعليم

التعليم العام
ن.ج/ر.ا

Bethlehem P.O.Box (168) : بيت لحم ص ب - Fax (02-2744392) فاكس - Tel (02-2741[illegible]) هاتف

بسم الله الرحمن الرحيم

State of Palestine

Ministry of Education & Higher Education

Directorate of Education & H. E. \Bethlehem

دولة فلسطين

وزارة التربية والتعليم العالي

مديرية التربية والتعليم العالي/بيت لحم

الرقم: 3178 /1/3

التاريخ: 2018.09.12م

الموافق: 2 محرم 1440هـ

مديري ومديرات المدارس الحكومية المحترمين

تحية طيبة وبعد،،،

الموضوع: تسهيل مهمة

نهديكم أطيب تحية، ونرجو من حضراتكم تسهيل مهمة الباحث **مجاهد علي موسى**، بإجراء دراسة بعنوان: Enhancing 4th and 5th grades inductive reasoning thinking skill at Palestinian schools، خلال الفصل الدراسي الأول من العام الدراسي 2019/2018م، دون أن يؤثر ذلك على سير العملية التعليمية.

مع الاحترام

أ. سامي كامل مروّة

مدير التربية والتعليم العالي

دولة فلسطين - وزارة التربية والتعليم العالي - مديرية التربية والتعليم العالي - بيت لحم

التعليم العام

ن.ح/ر.أ

Bethlehem P.O.Box (168) : بيت لحم ص.ب - Fax (02-2744392) فاكس - Tel (02-274....) هاتف

بسم الله الرحمن الرحيم

State of Palestine
Ministry of Education & Higher Education
Directorate of Education\Bethlehem

دولة فلسطين
وزارة التربيـة والتعليـم العالي
مديريـة التربيـة والتعليـم/بيت لحـم

الرقم: 3 / 1 / 2300
التاريخ: 2017.08.27م
الموافق: 5 ذو الحجة 1438 هـ

مديري ومديرات المدارس الحكومية المحترمين

تحية طيبة وبعد،،،

الموضوع: تسهيل مهمة

لا مانع من تسهيل مهمة الطالب **مجاهد علي موسى**، والسماح له بتطبيق دراسته بعنوان: **"Enhancing first graders thinking skills in Palestine"**، خلال الفصل الدراسي الأول للعام الدراسي 2018/2017م، على أن لا يؤثر ذلك على سير العملية التعليمية، علما بأن المعلومات لن تستخدم إلا لأغراض البحث العلمي.

مع الاحترام

أ. سامي كامل مروّة

مدير التربية والتعليم

دولة فلسطين
وزارة التربية والتعليم العالي
مديرية التربية والتعليم - بيت لحم

التعليم العام
ن.ح/ ر.أ

Bethlehem P.O.Box (168) : بيت لحم ص.ب - Fax (02-2744392) فاكس - Tel (02-2741271/2) هاتف

10.2 Appendix B: Permission letter to conduct research in the Ministry of Education

Doctoral School of Education

UNIVERSITY OF SZEGED

Petőfi sgt. 30-34., H-6722 Szeged, Hungary

Tel.: (+36-62)544163, 544032; Fax: (+36-62)420034

To: Ministry of Education and Higher Education

Ramallah, State of Palestine

Through: Directorate of Education/ Bethlehem

Cc: All principals of elementary and primary Schools with functional computer laboratories (internet connectivity)

Subject: permission to conduct research study at elementary and primary schools in Bethlehem in Palestine

Dear Mr. Mroah,

I am Mojahed Ali Mousa, passport No (3819597), ID No (850242611), a teacher at Directorate of Education/ Bethlehem (with annual vacation for study abroad) and currently a full-time PhD (Educational Science) student at the Doctoral School of Education of the University of Szeged in Hungary.

As part of my studies, my supervisor Professor Dr. Molnár Gyöngyvér and I would like to carry out a follow up research study on enhancing 4th and 5th graders inductive reasoning thinking skill at Palestinian schools.

The main goal of this study is to develop pupils' ability of inductive reasoning. This study is a follow up to the previous one which aimed to assess pupils' thinking skills, mainly of first graders (2nd and 3rd) through using an online diagnostic assessment.

The development of the Palestinian system regarding implementing technology in education presents the importance of this study. In addition, no studies have been established in inductive reasoning of school children in Palestine so far, therefore, this research will be considered as a valuable source.

We are looking forward to hearing from you.

Yours sincerely

Mojahed Mousa	Prof. Dr. Molnár Gyöngyvér
PhD student	Supervisor

10.3 Appendix C: Background information questionnaire

دائرة الدكتوراة في التعليمDoctoral School of Education

جامعة سجد UNIVERSITY OF SZEGED

Petőfi sgt. 30-34., H-6722 Szeged, Hungary

Tel.: (+36-62)544163, 544032; Fax: (+36-62)420034

إستبيان

عنوان البحث: تحسين مهارات التفكير لدى طلاب المرحلة الإبتدائية

هدف البحث: يهدف هذا البحث نشكل أساسي إلى تطوير وتنمية مهارات التفكير لدى طلاب المرحلة الإبتدائية وذلك عن طريق إستخدام التقييم التشخيصي عبر برنامج تعليمي محوسب يدعى edia . ويسعى الباحث إلى جمع عدد من البيانات والهدف المراد من هذه البيانات هو معرفة إذا كان هنالك تأثير للمتغيرات على مستوى تحصيل الطالب.

الباحث: أنا الباحث مجاهد علي موسى طالب بجامعة سجد في هنغاريا أسعى لنيل درجة الدكتوراة في علم التربية وهذا البحث هو جزء من رسالة الدكتوراة.

أرجو تعبئة هذا الإستبيان القصير بشكل صريح وموضوعي ويتعهد الباحث بالإبقاء على سريه هذه البيانات والتي ستستخدم فقط في مجال البحث العلمي.

- إسم الطالب: ....
- الصف: ....
- الجنس: ذكر – أنثى
- عدد ساعات إستخدام الطالب للحاسوب في اليوم (داخل البيت وخارجة)؟............................
- الدرجة العلمية للأم (إن وجدت) ....
- مهنة الأم.................................
- ضع دائرة: الدخل الشهري للأسرة: متدني – متوسط – جيد

شكراً لحسن تعاونكم

10.4 Appendix D: The questionnaire at the end of the training for the experimental group participants (see table 6.8. for the English version of the questionnaire).

هل أعجبتك اللعبة ؟

إختر الإجابة المناسبة:

- أعجبتني كثيرا
- أعجبتني، كانت جيدة
- أعجبتني ولم تعجبني في نفس الوقت
- لم تعجبني
- لم تعجبني كليا

التالي

كيف كان شعورك أثناء اللعب؟

إختر الإجابة المناسبة:

- جيد جدا
- جيد
- جيد أحيانا وأحيانا أخرى لا
- سيء
- سيء جدا

التالي

هل شعرت بالملل أثناء اللعب؟
إختر الإجابة المناسبة:

- دائما
- غالبا
- أحيانا
- نادرا
- قطعا

التالي

هل ستلعب هذه اللعبة مره أخرى إن سمح لك ؟
إختر الإجابة المناسبة:

- بالتأكيد
- غالبا
- ربما
- ربما لا
- بالتأكيد لا

التالي

هل ستلعب هذه اللعبة في البيت إن سمح لك ؟

إختر الإجابة المناسبة:

- بالتأكيد
- غالبا
- ربما
- ربما لا
- بالتأكيد لا

التالي

ما هي درجة صعوبة المهام ؟

إختر الإجابة المناسبة:

- كانت صعبة نوعا ما
- كانت صعبة
- كانت صعبة أحيانا وأحيانا أخرى لا
- لم تكن صعبة
- لم تكن صعبة إطلاقا

التالي

هل ساعدك نص المهمة عندما لم تكن تعرف الإجابة فورا؟

إختر الإجابة المناسبة:

- ساعدني كثيرا
- ساعدني
- أحيانا ساعدني وأحيانا أخرى لا
- لم يساعدني
- لم يساعدني إطلاقا

التالي

هل يسمح لك بإستخدام الحاسوب في البيت؟

إختر الإجابة المناسبة:

- نعم، يسمح لي بإستخدام الحاسوب في البيت
- لا يسمح لي بإستخدام الحاسوب في البيت

التالي