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

THE CONTEXT AND STRUCTURE OF THE DISSERTATION

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. The dissertation consists of seven main 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.

THEORETICAL BACKGROUND

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; Bridgeman, 2010; Csapó et al., 2012). As a result, a significant change in the effectiveness of assessments can be detected (Molnár & Csapó, 2019b).

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) (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).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).

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).

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). 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).

Empirical studies have indicated that reasoning 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).

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 (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.

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).

METHODS OF THE EMPICAL STUDIES

The samples were selected from schools in the directorate of Bethlehem city in Palestine (see Table 1). 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. Well-trained teachers helped in the data collection and supervised the whole training process. 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.

Table 1. The samples of the studies and the instruments

Study	Grade	Samples	Instrument	Method	Description
Study 1	2 & 3	Grades = 2 & 3 *N* = 57	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	Grades = 2, 3 & 4 *N* = 193	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	Grades = 4 & 5 *N* = 248	IR (Molnár & Csapó, 2019a).	eDia platform	IR tasks - domain general and culture-free content reasoning tasks. Numerical items included.
Study 4	4 & 5	Grades = 4 & 5 *N* = 236	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.

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). 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.

RESEARCH AIMS AND EMPIRICAL STUDIES

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.

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

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).

The applicability of the IR test and its subtests were supported by the two-dimensional IRT analyses. The items were generally well-matched to the sample, but there were items which were relatively hard for the pupils 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.

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

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.

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.

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.

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

The internal consistencies 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.

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.

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

No significant difference was found between the performance of the experimental group and that of the control group prior to the experiment (*Mcont* = 33.4%, *SDcont* = 15.0; *Mexp* = 35.0%, *SDexp* = 13.5, t = 1.3, p = .18). While no significant development could be detected in the case of the control group (*Mcont* = 34.0%, *SDcont* = 14.1), the experimental group significantly outperformed the control group by more than one standard deviation (*Mexp* = 58.6%, *SDexp* = 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.

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 (see Figure 1.).

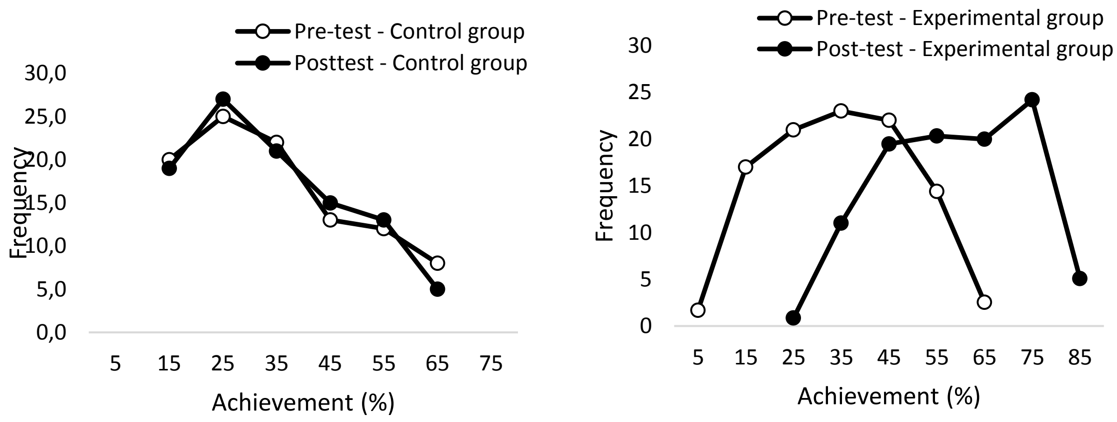

Figure 1. Distribution curves for the control and experimental groups in the pre- and post-test

The group-level results above are supported by the pupil-level analyses, illustrated in Figure 2., where the performances in the pre- and post-tests are projected onto each other. 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. Therefore, the training proved to be gender-fair.

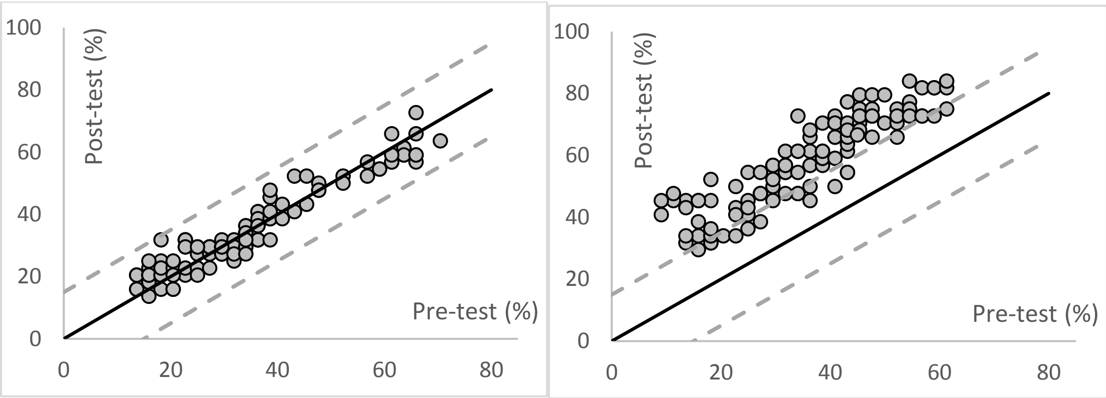

Figure 2. Pupil-level changes in achievement from pre-test to post-test in both the control and experimental groups

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.

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). The developmental results for the experimental group show that the training had the same effect on all the pupils, independent of their school achievement and initial IR skill level.

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. 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.

GENERAL 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. This study has successfully discovered several pieces of valuable information.

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. We confirmed what we found in the analysis of the results, that the Palestinian school system supports the explicit development and evaluation of inductive reasoning. We reached some interesting research results out of the training program 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 (de Koning et al., 2002; Molnár, 2011), and (3) the lack of explicit fostering of inductive reasoning in school (de Konig, 2000; Molnár, 2011).

The study proved that the assessment or the training were not related to the gender which means it had a similar effect on boys and girls. 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). In addition, the study provided evidence that the mother's educational and occupational backgrounds, and the school achievement can influence the achievement of the participating pupils.

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.

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).

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.

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

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.

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.

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.

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.

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

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. (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., 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

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

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

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

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

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

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.

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.

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.

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. (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. (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., 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. & 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.

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

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

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

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.

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

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

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

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.

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

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.

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

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

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.

PUBLICATIONS AND CONFERENCE PRESENTATIONS RELATED TO THE TOPIC OF 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.