

The Feasibility of Computer-based Testing in Palestine Among Lower Primary School Students: Assessing Mouse Skills and Inductive Reasoning

Mojahed Mousa

PhD candidate at University of Szeged

Institute of Education, address: H-6722 Szeged, Petőfi sgt. 30-34., Hungary

E-mail: mojmousa@gmail.com

Gyöngyvér Molnár

Director of the Institute of Education at University of Szeged

Address: H-6722 Szeged, Petőfi sgt. 30-34

E-mail: gymolnar@edpsy.u-szeged.hu

Received: Jan. 25, 2019

Accepted: April 17, 2019

Published: May 1, 2019

doi:10.5296/jse.v9i2.14517

URL: <https://doi.org/10.5296/jse.v9i2.14517>

This study was funded by OTKA K115497.

Abstract

This study introduces and explores the potential of using computer-based testing at Palestinian schools. It investigates the developmental level of mouse skills among year 2 and 3 students and tests the applicability of an online test measuring pupils' inductive reasoning. The sample for the study was drawn from year 2 (N=28) and 3 (N=29) students in Palestinian primary schools (Mean_age=7.5; SD=.50). The instruments consisted of 28 figural items for the mouse usage test and 36 figural items for the inductive reasoning test. The eDia system was used to collect the data. The reliability coefficient of the mouse usage and inductive reasoning tests were .75 and .915, respectively. Results showed that the mouse usage test was easy (M=90.53% SD=9.67%), while the inductive reasoning test was moderately difficult for the students at this age (M=44.15, SD=24.28). The frequency of computer usage did not influence test achievements. There were no gender level differences detectable in students'

mouse skills. Girls achieved significantly higher in the inductive reasoning test ($M_{girls}=50.88$, $M_{boys}=37.67$; $t=-2,15$, $p<.05$). The results proved that computer-based testing can be used in Palestine and it can work effectively even when it comes to early age pupils.

Keywords: Computer-based testing, inductive reasoning, assessment, thinking skills, technology

1. Introduction

The usage of technology in recent years is not just limited to entertainment, project designs, analysis, but it is also included in education like e-learning or edutainment. These new methods take the advantage from the traditional ones which basically based on “drill and practice” (Brom et al., 2009). The advantages of digital-based learning in enhancing students’ involvement and creating a combination between learning and entertainment attracted the attention of educators and researchers (Prensky, 2007), by that, computerised testing becomes an interesting area of educational evaluation to be developed rapidly (Csapó et al., 2012).

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 Year 2 and 3 students and tests the applicability of an online test measuring pupils’ inductive reasoning using the ICT facilities of the participating schools.

2. Background information

2.1 *The transition 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 testing, a number of questions arise (e.g. equivalence issue, mode-effect, validity, level of technology usage, infrastructure, security).

Several studies have been conducted about the shifting from paper-based 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 students’ 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 students achievement in a technology-based environment. By time, the differences between PP and CB test performances are well documented. The majority 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 students as computers become more broadly accessible at schools (Way et al., 2006).

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, 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). Thus computer-based testing is an “innovative” approach to assessment (Thurlow et al., 2010), however, there are still challenges for using it

by young learners (Csapó, Molnár, & Nagy, 2014; Carson, Gillon, & Boustead, 2011; Choi & Tinkler, 2002).

2.2 Mouse and Keyboarding Skills

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 of a 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 (2015) explored the potential of using computer-based tests in regular educational practice for the assessment of pupils in the beginning of 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 only typing 1 to 5 numbers or letters finally, drag-and-drop 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 it 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, 2015, p. 118).

2.3 Inductive reasoning

Inductive reasoning is considered as a general thinking skill (Molnár et al., 2013). It is a basic component of the thinking processes (Klauer & Phye, 2008; Molnár, 2011). Its connection to higher order thinking skills is strong (Molnár et al., 2013; Schubert 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).

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). A classical interpretation of inductive reasoning defines it as the process of moving from the specific to the general (Sandberg & McCullough, 2010). We adopt Klauer's interpretation (1993) and define inductive reasoning 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" (Molnár, 2011: 92). Six different processes can be derived based on this approach: generalization, discrimination, cross-classification, recognizing relations, discriminating 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).

Findings from prior research have highlighted the importance of inductive reasoning in knowledge acquisition and application (Goldman & Pellegrino, 1982; Bisanz, Bisanz, & Korpan, 1994; Hamers, De Koning, & Sijtsma, 2000; Klauer, 1990, 1996; Pellegrino & Glaser, 1982), in school learning for example second language (Csapó & Nikolov, 2009). IR have been defined as a component skill of problem solving (Wu & Molnár, 2018). It helps students to acquire a deeper understanding of the subject matter in the classroom (Molnár, 2011). It also helps researchers to set new hypotheses to make predictions about specific issue (Hayes et al., 2010). Moreover, the generalisation feature in inductive reasoning make it a handy method. However, the relations between inductive reasoning and intelligence are strong, thus they are closely connected to each other (Klauer et al., 2002; Csapó, 1997). Findings suggests that its' fostering is not an integral part of school curricula (de Koning, 2000), however it can be developed effectively (Klauer & Phye, 2008; Molnár, 2011). Therefore, it has been suggested by several scholars to include thinking skills activities at the school curriculum (Molnár, 2011; de Konig, 2000; Resnick, 1987).

2.4 The Palestinian education system

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 shows that around eighty percent of the governmental school are well equipped by computer laboratories and more than two-third of these labs are connected to the internet (PMEHE, 2018).

3. Aims and research questions

In the present study, we explore the possibilities of the application of computer-based assessment in regular Palestinian educational practice at the early stages of schooling. For this purpose, we apply online tests that already have established psychometric characteristics; we adapt them to Arabic in both sense: language and directions and monitor whether computer-based assessment is applicable at the Palestinian context or not.

More specifically, we aim to answer the following research questions:

1. Is computer-based assessment feasible and applicable among young learner in the Palestinian schools?
2. Which background factors influence the applicability of computer-based testing among 2nd and 3rd year Palestinian students?
3. Are there gender differences regarding students mouse and keyboarding skills and their achievement on the inductive reasoning test?

4. Is there any relationship between students' school achievement and the developmental level of IR skills?

4. Methods

4.1 Participants

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

4.2 Instruments

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

The mouse usage 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 prerecorded voice.

Pupils' basic computer skills, such as keyboarding and mouse skills is measured in the first test. The test investigates the nature and developmental level of mouse skills by testing two processes: clicking and drag and drop. Students 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 1). The size and amount of the objects they had to click on it or drag-and-drop was changing in the test. In the first tasks they had to click and drag 1-3 big elements to big places, while at the end of the test they had to cope with clicking and dragging more than three small elements. Students 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. Students' 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 1. A sample item from the mouse usage test

The inductive reasoning test comprised two subtests: figural series and figural analogies (see Figure 2). In case of figural series, students are 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 figural analogy, students are required to find out the missing shape which fits into the yellow box regarding the result of changes in the other two boxes.

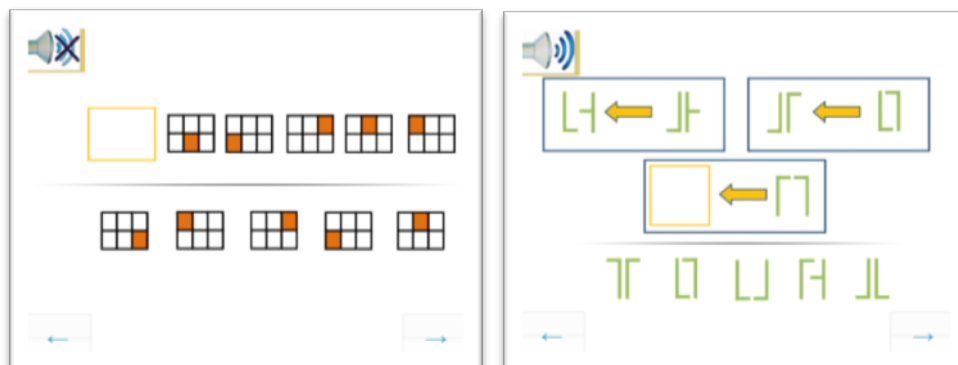


Figure 2. Series and analogy items of the inductive reasoning test

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 3). The instructions have been translated into Arabic language, recorded and attached to the system, thus they were given online using headsets. Automatic scoring was used and instant feedback was provided at the end of the tests. In this case students had only one single attempt per tasks, thus students' 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").

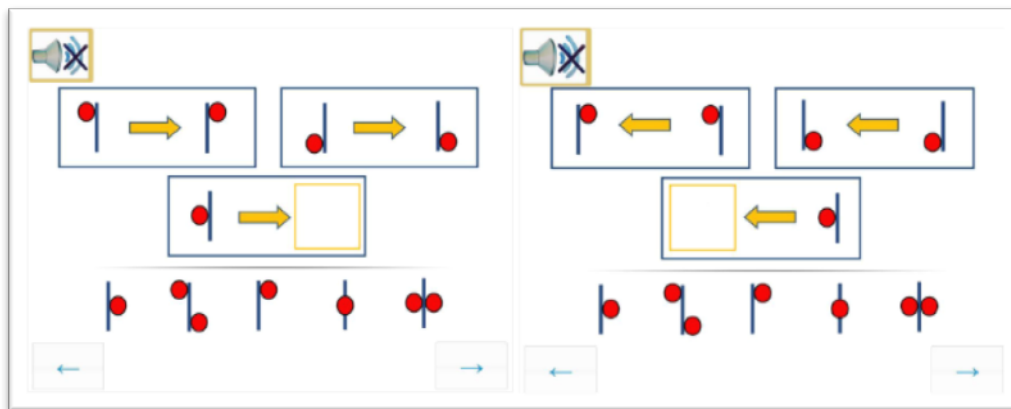


Figure 3. Sample item from the inductive reasoning test (left: original Hungarian version, right: translated Arabic version)

4.3. Procedures

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, students were provided with instructions, in which they could have learnt how to use the program: (1) at the top of the screen, a yellow bar indicates how far along they are in the test; (2) they must click on the speaker to be able to listen again to the task instructions; (3) to move on to the next task, they must 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 receive immediate visual feedback with 1 to 10 balloons, where the number of balloons is 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 at 50% the probability of success 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.

5. Results

5.1 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 usage test and .92 for the inductive reasoning test. The subtest level analyses regarding IR test reveal that the subtest level results are also generalizable

(Cronbach's $\alpha=.849, .853$; EAP/PV Reliability $=.865, .865$), however, they show an increased probability of measurement errors. The mouse usage test was generally easy ($M=90.53\%$ $SD=9.67\%$) for pupils, especially for grade 3 students ($M=97.0\%$, $SD=2.87$), 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 students at this age ($M=43.46$, $SD=23.7$).

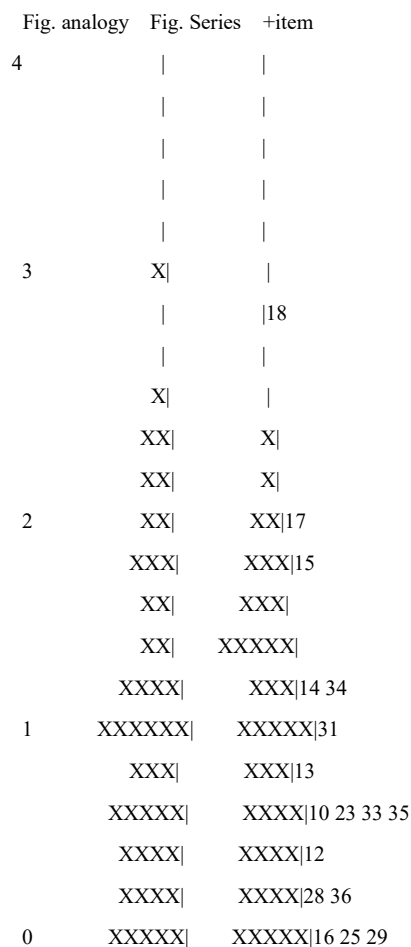
The learning effect regarding the basic computer skills is supported by the result. Students achievement proved to be significantly lower ($M=84.46$, $SD=10.89$) and the differences between students significantly higher without task-level feedback and second trial ($t=-7.65$, $p<.001$). Independent from the required procedure tasks requiring the same or similar operations than items somewhere previously in the test were significantly easier than items consisting the same operation for the first time in the test. These results indicate that students 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 students computer skills into account, CBA can be used for measuring students' performance at this early school age. Students' mouse skills are adequate to answer computer-based tests requiring clicking and drag-and-dropping of big and smaller object or objects. Therefore, the relative lower achievement (Table 1) on the inductive reasoning test is not caused by students' mouse skills, but their level of inductive reasoning skills. There was no significant correlation between students mouse skills and achievement on the IR test ($r=.215$, $p>.05$).

The test comprised of two subtests measuring inductive reasoning by figural series and figural analogy. Students 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. Students, whose achievement was high on figural series, tend to achieve high on the figural analogy tasks too ($r=0.89$, $p<.001$) and on the other way round, students, whose achievement was lower on the first subtest, achieved lower on the second one too.

Table 1. Descriptive statistics and reliability indexes regarding the inductive reasoning test

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 are supported by the two-dimensional IRT analyses too. Figure 4 shows the match between the item difficulty distribution and the distribution of students' Rasch-scaled achievement estimates for inductive reasoning. The two-left panel, headed Fig. Analogies and Fig. Series, show the distribution of students' achievement on the two areas of IR, respectively. Students at the top end of these distributions have higher achievement estimates than students 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 on the areas. The units of the vertical scale are not scaled, because they are in these Figures arbitrary. In Figure 4 every 'x' represents 0.5 students.



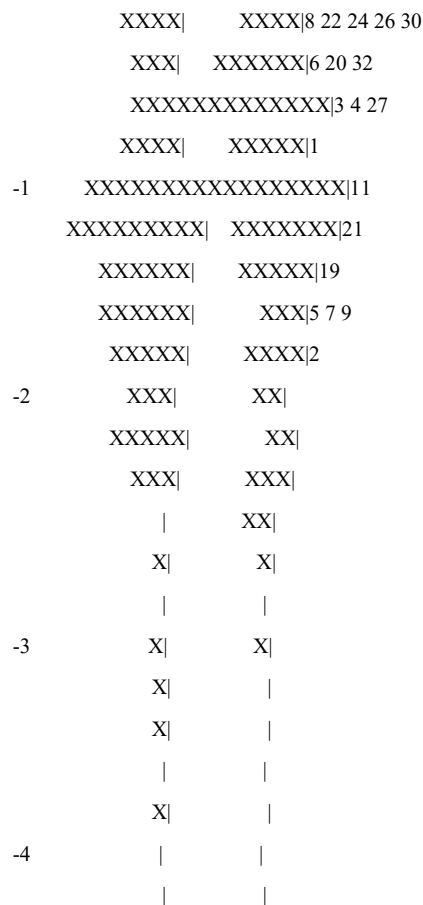


Figure 4. The two-dimensional item/person map of the inductive reasoning test

The items are generally well-matched to the sample ('x' and number are almost parallel) but there are items which are relatively hard for the students (e.g. item 18) and there are more easy items missing from the IR test. Students behaved similar in both of the 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 Palestinian educational context even at the early age of schooling. The reliability of assessments proved to be good and generally children do not have difficulties handling the computerized tests. The difficulty of the IR test fit the ability level of the sample well both on test and subtest level and students' achievement on it was independent from their level of mouse skills.

5.2 Factors influencing the applicability of computer-based assessment by pupils

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

5.3 Gender differences

There were no gender differences detectable in the mouse skill test ($M_{\text{male}}=91.51\%$, $SD=8.25$; $M_{\text{female}}=89.28\%$, $SD=11.29$; $t=.83$, $p>.05$). Thus eventual gender differences on the IR test could have not be caused by the differences in the mouse skills.

In the inductive reasoning test girls achieved significantly higher ($M_{\text{girls}}=50.88$, $M_{\text{boys}}=37.67$; $t=-2.15$, $p<.05$) 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 2).

Table 2. Grade and gender-level differences in students' 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		

5.3 Relationship between students' school achievement and the developmental level of IR skills

Students with higher achievement at school achieved significantly higher ($M_{\text{school_advanced}}=60.49$, $M_{\text{school_average}}=44.3$; $M_{\text{school_low}}=26.46$, $F=13.82$, $p<.001$; $r=.58$, $p<.001$) on the IR test. Students with higher school achievement proved to be more developed in inductive reasoning skills, thus the development and evaluation of students IR skills are hidden embedded in the Palestine school curricula and evaluation process.

6. 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 a number of challenges and questions concerning the validity of results.

We confirmed research results from the literature (Molnár & Pásztor, 2015) that computer-based testing and training can be used at 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 students' basic ICT skills and test achievement, thus students 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.

A number of studies have been conducted in different countries 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 reported strong correlations between reasoning skills and successful learning of several school subjects, for example, second languages (Csapó & Nikolov, 2009). The finding of the present study suggest that the Palestinian school system supports the explicit development and evaluation of inductive reasoning. Students with higher school achievement proved to be more developed, students with lower school achievement proved to be less developed in inductive reasoning skills. Thus, the development and evaluation of students’ inductive reasoning skills must be embedded in the Palestine school curricula and evaluation process.

7. Limitations

Due to the context of the study, the first piloting phase of a larger project, there were smaller sample size available for the analyses. A further limitation of the present study is that only students’ 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.

References

- Barnes, S. K. (2010). Using computer-based testing with young children. 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
- Bisanz, J., Bisanz, G., & Korpan, C. A. (1994). Inductive reasoning. In R. Sternberg (Ed.), *Thinking and problem solving*. San Diego: Academic Press. <https://doi.org/10.1016/B978-0-08-057299-4.50012-8>
- 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. <https://doi.org/10.1007/s00530-009-0174-0>
- 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>
- 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.
- Clariana, R. & Wallace, P. (2002). Paper based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(1), 593–602. <https://doi.org/10.1111/1467-8535.00294>

- 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., & 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., Lőrincz, A. & Molnár, G. (2012). Innovative assessment technologies in educational games designed for young students. *Assessment in Game-Based Learning, 235-254*. https://doi.org/10.1007/978-1-4614-3546-4_13
- 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. & 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* (113–118). Luxemburg, Belgium: Office for Official Publications of the European Communities.
- de Konig, E. (2000). *Inductive Reasoning in Primary Education: Measurement, Teaching, Transfer*. Zeist, 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>
- 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.
- 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](https://doi.org/10.1016/0010-0285(83)90002-6)
- 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>
- 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>

- 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>
- 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 [Has inductive Thinking a positive effect on complex problem solving?]. *Zeitschrift für Experimentelle Psychologie*, 43(1), 361–366 (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. J. (1993). *Denktraining für Jugendliche*. Göttingen: Hogrefe.
- Klauer, K. J., & Phye, G. D. (2008): Inductive reasoning: A training approach. *Review of Educational Research*, 78(1), 85–123. <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>
- Molnár, G. & Pásztor, A. (2015). Feasibility of computer-based assessment at the initial stage of formal schooling: the developmental level of keyboarding and mouse skills in Year One. In *Conference on Educational Assessment*, Hungary, April 23-25, 2015, p. 118.
- Molnár, G. (2011). Playful fostering of 6- to 8-year-old students' inductive reasoning, *Thinking Skills and Creativity*, 6(2), 91-99. <https://doi.org/10.1016/j.tsc.2011.05.002>
- Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, 9, 35-45. <https://doi.org/10.1016/j.tsc.2013.03.002>
- 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>
- 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.
- PMEHE (2018). *Educational Statistics 2016/2017*. Palestine Ministry of Education and Higher Education, Ramallah. Retrieved on May 2018, from <http://www.mohe.pna.ps/services/statistics>
- Prensky, M. (2007). *Digital game-based learning*. Saint Paul, MN: Paragon House.

- Resnick, L. (1987). *Education and learning to think*. Washington, DC: National Academy Press.
- 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.
- Scheuermann, F., & Pereira, A. G. (2008). Towards a research agenda on Computer-based Assessment. *Challenges and needs for European Educational Measurement*. Luxembourg.
- Schubert, T., Astle, D. & Van Der Molen, M. (2012). Computerized training of non-verbal reasoning and working memory in children with intellectual disability. *Training-induced cognitive and neural plasticity*.
- Shraim, K. (2018) Palestine (West Bank and Gaza Strip). In: Weber A., Hamlaoui S. (eds) *E-Learning in the Middle East and North Africa (MENA) Region*. Springer, Cham, pp 309-332. https://doi.org/10.1007/978-3-319-68999-9_14
- Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, 25, 231-280. <https://doi.org/10.1006/cogp.1993.1006>
- 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*. The Open University, Heerlen, The Netherlands. 20(1), 483-490. <https://doi.org/10.1006/ceps.1995.1036>
- 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.
- Wu, H., & Molnár G. (2018). 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>