University of Szeged
Faculty of Arts
Doctoral School of Education
Program of Learning and Instruction

Assessing Namibian Students' Abilities in Scientific Reasoning, Scientific Inquiry and Inductive Reasoning Skills

Summary of the PhD Dissertation

Linus Kambeyo

Supervisor: Prof. Dr. Benő Csapó

Szeged, Hungary, 2018

THE CONTEXT AND STRUCTURE OF THE DISSERTATION

The effectiveness of basic education is often evaluated through low-stakes educational assessment studies both at a national and an international level (Vainikainen, 2014). Such assessments include end-of-year tests and examinations. In Namibia, the only national assessment is that which assesses specific subject content at the end of the primary phase, junior secondary phase and secondary phase. It is an unfortunate fact that students continue to perform poorly in these subject content-based tests. It is also vital to note that Namibian students do not participate in any international assessment programs such as Programs for International Student Assessment (PISA), Trends in International Mathematics and Science Studies (TIMSS) or The Progress in International Reading Literacy Study (PIRLS). Moreover, an adequate level of general skills (inductive reasoning), scientific reasoning and scientific inquiry skills is necessary in order to develop and operate these competencies.

However, it has not yet been established at which level, in terms of international assessment, students from Namibia are situated, with regard to scientific reasoning and inquiry skills and inductive reasoning skills. Many educational experts recommend the need for developing and assessing thinking and reasoning skills in the everyday school context (Adey, Csapó, Demetriou, Hautamaki, & Shayer, 2007; Zimmerman, 2007). A developed country like Finland provides an example of a system in which the monitoring of the educational outcomes is based entirely on sample-based assessments, which do not normally have any consequences for the participating students at an individual level (Vainikanen, 2014).

In order to foster students' thinking and reasoning skills in the classroom context and monitor their development, reliable and valid, easy to use assessment tools have to be available for teachers to use. Because Namibian students have not been exposed to these kinds of tests before, no data exists regarding their scientific reasoning, scientific inquiry and inductive reasoning skills. Moreover, the advancement of technology-based assessment (TBA) is gaining momentum worldwide. We therefore wished to explore the viability of technology-based assessment in Namibia as well.

To date many studies have been conducted on assessing students' scientific reasoning skills, general thinking skills (inductive reasoning), and scientific inquiry skills, mostly in developed countries (Bao et al., 2009; Csapó, 1997; Han, 2013; Mayer et al., 2014; Klauer, 1999, 2002; Wenning, 2007). Results from these studies have shown that assessing these skills helps improve the quality of teaching and learning in schools, especially in the Science, Technology, Engineering and Mathematics (STEM) fields. It also improves the education system in general by enhancing the thinking and reasoning skills of the students. Informed by the literature, the current study about the assessment of scientific reasoning, inductive reasoning and scientific inquiry skills was deemed necessary to act as a stepping-stone toward further research in these areas in Namibia. The study attempts to ascertain the thinking and reasoning abilities of Namibian students. To achieve this, four sub-studies were carried out using the cross-sectional research paradigm.

The dissertation consists of six chapters. Chapter One introduces the study, outlining its context and its motivation. Chapter Two provides a review of literature relevant to the study. The literature reviewed is on the main three main constructs studied, i.e., Scientific reasoning (SR), Scientific inquiry (SI) and Inductive reasoning (IR). The general consensus from the literature is that students need to be taught not only STEM content, but also the skills needed for survival in the 21st century. Countries and schools are encouraged to carry out low-stakes assessments at school and at national level. A brief synopsis of the importance of technology-based assessment is also provided in this chapter. Chapter Three furnishes the research aims, questions and hypotheses guiding the study.

Chapter Four discusses the methodologies employed to frame, plan and carry out this research study. It also describes the study's research paradigm: a cross sectional quantitative approach. The research site and sampling procedures are discussed in detail. Data gathering techniques are described, as well as how the data was analyzed and validated.

Chapter Five presents the analyzed data from the four complementary studies conducted for this project, which explored the possibilities of online assessment, examined students' knowledge and thinking skills, and investigated the relationship between reasoning skills in a science context and the motivation to learn science in the Oshana region, Namibia.

Each study is discussed in detail according to the research sub-questions guiding that specific sub-study. Sub-study 1 used technology-based assessment to assess the 5th and 7th grade students in cognitive tasks such as scientific reasoning and inductive reasoning.

The second sub-study was then carried out with a much older sample (9th and 11th grade students). The purpose of this study was to 1) explore the possibilities and feasibilities of an online assessment of scientific inquiry skills; 2) examine the psychometric properties of the test; and 3) ascertain the ability level of the 9th and 11th graders' scientific inquiry skills.

Because of the results of the second sub-study, in which the students' performance improved compared to the 5th and 7th graders, a decision was made to assess the 10th and 12th grades using paper and pencil. This was also informed by the fact that the schools that were selected to take part in the research lacked sufficient ICT equipment and reliable internet connectivity. In addition, these are the also the grades that write national exit examinations. It became necessary to switch to pencil and paper methods in order to assess all three skill sets, scientific reasoning, scientific inquiry, and general thinking such as inductive reasoning. The relationships among scientific reasoning, inductive reasoning and scientific inquiry were explored.

The last sub-study, sub-study 4, focused on students' reading comprehension skills and their impact on students' performance in IR and SR. The students' socio-economic status was also explored to ascertain whether it affected their performance. The reason for bringing in the reading comprehension text was that so many students performed below average in the other tests, and it was thought that they might not have an adequate level of comprehension of the medium of instruction, English. The aim of the sub-study was to find out if SES affects reading skills, inductive reasoning and scientific reasoning achievement. We also wished to explore if reading skills influenced inductive and scientific reasoning skills.

Finally, Chapter Six concludes the study by providing a summary of the findings of the study, making certain recommendations arising from these findings, describing the limitations and implications of the study, and suggesting avenues for future research.

THEORETICAL BACKGROUND

Research on scientific reasoning is rooted in early studies of the cognitive development of ‘formal reasoning’ and ‘critical thinking’ (Hawkins & Pea, 1987). Traditionally, the Piagetian clinical interview is used to assess students’ formal reasoning abilities. It has been argued that the current style of the content-rich STEM education, even when carried out at a rigorous level, has little impact on the development of students’ scientific reasoning abilities (Bao, et al., 2009). The findings of the comparative study by Bao et al. (2009) of American and Chinese students indicate that it is not what we teach, but rather how we teach, that makes a difference to students’ higher-order abilities in scientific reasoning. Students ideally need to develop both content knowledge and transferable reasoning skills (Bao et al., 2009). The onus is upon researchers and educators to invest more time in the development of a balanced method of education that incorporates more inquiry-based learning that targeting both goals.

Other studies have indicated that scientific reasoning is critical in enabling the successful management of real-world situations in professions beyond the classroom (Han, 2013). For example, in the K-12 curriculum in the USA, the development of scientific reasoning skills has been shown to have a long-term impact on students’ academic achievement (Adey & Shayer, 1994). Positive correlations between students’ scientific reasoning abilities and measures of students’ gains in learning science content have been reported (Coletta & Phillips, 2005), and reasoning ability has been shown to be a better predictor of success in Biology courses than prior Biology courses (Lawson, 2000).

Scientific inquiry (SI) has always been an integral part of scientific literacy (Bybee, 2009) and a long-standing area of research and discussion in science education (Fenichel & Schweingruber, 2010; Yeh, Jen, & Hsu, 2012). Scientific inquiry is seen as a problem-solving activity (Klahr, 2000). It can also be viewed as a circular process, in which research questions and hypotheses are formulated, investigations are planned and carried out, and evidence is evaluated with regard to the hypotheses and the underlying theory (Mayer, 2007; Zimmerman, 2005). The literature emphasizes that scientific inquiry leads to knowledge and understanding of the natural and synthetic world through direct interaction with the world, and through the generation and collection of data for use as evidence in supporting explanations of phenomena and events (Harlen, 2013). Teachers are therefore encouraged to adopt and make use of the approaches emphasized in inquiry-based science education (IBSE) if they are to instill in children a culture of exploring, experimentation, investigation and explanation.

The third skill measured was the domain of general reasoning skill, inductive reasoning (IR). Inductive reasoning is considered a basic component of thinking, and it is one of the most broadly studied procedures of cognition (Csapó, 1997). The inductive method, or teaching by examples, is one of the oldest methods of instruction. Induction can also be defined as the process whereby regularities or order are detected and, inversely, whereby apparent regularities or generalizations are disproved or falsified. This is achieved by finding out, for instance, that all swans observed so far are white or, on the contrary, that at least one single swan has another colour. Put differently, one can state that the process of induction takes place by detecting commonalities through a process of comparing.

What is common to all the literature reviewed about inductive reasoning is that induction is one of the fundamental thinking processes and plays a central role in human cognition, knowledge acquisition and the transfer of knowledge. The educational relevance of the skill is therefore well grounded. From an educational perspective, Klauer and colleagues' analysis of inductive reasoning offers a well-structured and detailed theory and system (Klauer, 1989, 1990, 1996, 1997; Klauer & Phye, 1994; Klauer et al., 2002). The assessment items designed and used in this study were based on this approach and are best understood in terms of its framework.

With regard to TBA, it is argued that technology significantly advances educational assessments in a number of dimensions (Csapó et al., 2012). For example, it was found to improve the efficiency of collecting and processing data; it enables the sophisticated analysis of the available data, supports decision-making and provides rapid feedback to participants and stakeholders alike. However, although TBA has proven to be an efficient way of doing things in some parts of the world, there are still challenges facing countries in the developing world such as Namibia – challenges such as the provision to schools of internet infrastructure, computers, tablets and all other technological equipment.

Most studies have not looked into the relationship between domain-specific skills, i.e. scientific reasoning and scientific inquiry skills, which alone makes this research study necessary. Moreover, most cognate studies have been carried out in the developed world, so there is no knowledge of students' abilities in these skills in a country like Namibia.

STUDY AIMS AND OBJECTIVES

The purpose of this study is to assess students' abilities in thinking and reasoning skills such as, scientific reasoning, scientific inquiry and inductive reasoning skills in Namibia**.** We also wish to explore the possibilities of an online assessment in Namibia. This study connects three rapidly developing areas of educational research and places them in the context of the development of the Namibian education system. First, improving the conditions of science education in order to attract more young people to the Science, Technology, Engineering, and Mathematics (STEM) professions is a goal of many countries as the supply of young professionals graduating in these areas does not meet the demand of modern economies. One of the main directions in this area is the quality of science education, especially its contribution to the improvement of students' higher order thinking skills. Second, educational assessment is receiving growing attention both in research and in practice. If certain psychological constructs are made measurable, it opens a path for conducting precise training experiments, while feedback provided by the assessment may orient practice. Third, testing is transferred to a technological basis making even more constructs measurable, while reducing the costs and timeframe of the assessments.

This study consists of four empirical sub studies (1) online assessment of scientific reasoning, inductive reasoning and motivation to learn science, the aims of this study were to explore the possibilities of online assessment, to investigate students' knowledge and reasoning skills based on log file analysis, and to explore the relationship between reasoning skills in a science context and the motivation to learn science in Oshana region, Namibia. (2) The possibilities of assessing students' scientific inquiry skills abilities using an online instrument, the aim of study 2, was to pilot the on-line instrument for the assessment of scientific inquiry skills of the 9^{th} and 11^{th} grade students. Primarily, the study wishes to construct a scale tapping different components of scientific inquiry to measure it broadly, economically and efficiently using an on-line assessment platform, Electronic Diagnostic Assessment (eDia). (3) Assessment of scientific reasoning, scientific inquiry and inductive reasoning with students' socioeconomic status. The relationship among the three main constructs (SR, SI, & IR) for this study was also explored, (4) and the forth supplementary study was conducted to further explore if English reading comprehension skills affect students' achievement in SR and IR. Generally, these four studies aim to assess students' abilities in thinking and reasoning skills such as, scientific reasoning, scientific inquiry and inductive reasoning skills in Namibia**.** We also wish to explore the possibilities of an online assessment in Namibia. Data collection procedures, timeline and steps are summarized in (table 1).

Table 1 The timeline and the research activities

Timeline	Research activities	Instruments	Samples
September 2015	• Contextualization of the research focus, writing and presenting research plan. • Development and translation of assessment items. • Literature search and reviews.		
June-July 2016	• Piloting the assessment instruments in the	• Scientific reasoning-(Korom et al., 2012; 2017) • Inductive reasoning -(Pásztor et al., 2017) • Science Motivation questionnaire-(Glynn et al., 2011) • eDia platform	• 5th & 7th grades, • N=616
September 2016	• Data analysis and presentation of the results.		
January 2017	• Assessing the students' through an online platform.	• Scientific inquiry skills test-(Nagy et al., 2015; Korom et al., 2017) • eDia platform	• 9th and 11th grades • N=118
February 2017	• Data analysis of the collected data in January 2017 and presentation of results to conferences and seminars. • Organizing final assessment tools for the large-scale assessments.		
June-August 2017	• Large scale assessment of Scientific reasoning, Scientific inquiry and Inductive reasoning skills in the. •	• Lawson Classroom Test of Scientific reasoning skills-(Lawson, 2000) • Scientific inquiry skills- (as above) • Inductive reasoning- (as above) • Paper and pencil methods.	• 10th and 12th grades • N=582
September 2017	• Data analysis continues and presentation of results in conferences and seminars.		
January/February 2018	• Supplementary assessment of the students Reading comprehension skills, Scientific reasoning and Inductive reasoning.	• Reading comprehension skills tests -(Csapó & Nikolov, 2009; 2018) • Scientific reasoning-(Lawson, 2000) • Inductive reasoning- (as above) • Paper and pencil methods used	• 8th grade students • N=250
March 2018	• Data analysis and results presentations.		

RESEARCH QUESTIONS AND HYPOTHESES

Research questions and hypotheses are divided according to the four sub studies that guided the research. More specifically, corresponding research questions and hypotheses for each sub study are as follow.

Research question for sub study 1

1. What are the psychometric properties of the SR and IR tests? Will the tests be reliable?
2. How well do grade five and seven students perform on scientific and inductive reasoning tests?
3. How does girl students' performance differ from boy students' in scientific and inductive reasoning?
4. How do the background variables (e.g. parents' level of education) influence their performance?
5. What are the relationships among the subtests of scientific reasoning skills?
6. What is the relationship between inductive and scientific reasoning, scientific reasoning and the motivation to learn science of the students?

Hypotheses for sub study 1

H_1. The psychometric properties of the tests are acceptable.

H_2. We expect the 7^{th} graders to perform significantly better than the 5^{th} graders.

H_3. Based on the literature, we expect no significant differences between genders.

H_4. In accordance with the literature, we hypothesized that students from high socio-economic backgrounds to significantly perform better that students from low economic backgrounds.

H_5. We hypothesized medium correlations between the subconstructs.

H_6. We expect strong correlation between SR and IR, and between SR and motivation to learn science (SMQII).

Research question for sub study 2

1. What are the psychometric properties of the scientific inquiry skills test?
2. How well do grade 9 and 11 students perform on the scientific inquiry skills?
3. Is there a significant difference in performance between genders on scientific inquiry test?
4. How do parents' levels of education influence students' performance?
5. What are the relationships among the subtests of the scientific inquiry skills test?

Hypotheses for sub study 2

H_7. We expect the psychometric properties to be acceptable.

H_8. We hypothesized significant achievement differences between the two grades

H_9. We expect no significant differences in achievement between genders.

H_{10}. Literature have emphasized that mothers' level of education influence students' achievement, therefore we expect high socio-economic status students to significantly perform better than low socio-economic status.

$H_{11.}$ We expect strong correlations between the subconstructs and the whole scale.

Research question for sub study 3

1. Are the psychometric properties of the tests acceptable in this sample?
2. How well do the 10^{th} and 12^{th} grade students perform on scientific reasoning, inductive reasoning and scientific inquiry skills? (What are the differences among the three tests within the grade?)
3. How do girls differ from boys on scientific reasoning, inductive reasoning and scientific inquiry skills?
4. How do the background variables (e.g. parents' level of education) influence their performance?
5. What is the relationship among the subtests of scientific reasoning, inductive reasoning and scientific inquiry skills?

Hypotheses for sub study 3

$H_{12.}$ We expect the psychometric properties to be acceptable.

$H_{13.}$ Significant differences in achievement between the two grades is anticipated.

$H_{14.}$ As per the literature, no significant differences between gender is hypothesized among the three constructs.

$H_{15.}$ Based on our previous research results, we hypothesized significant differences in achievements between low and high socio-economic students.

$H_{16.}$ Based on our previous results, strong correlations are expected between sub constructs and whole scales

Research question for sub study 4

1. How well do the 8^{th} graders perform in reading comprehension (RC), IR and SR?
2. What are the relationships between students' RC skills and their performance in IR and SR?
3. How does students' SES affect their performance in IR and SR?

Hypotheses for sub study 4

$H_{17.}$ Based on our previous research results, we hypothesized better achievements in the tests.

$H_{18.}$ We expect strong correlations between the three tests.

$H_{19.}$ As with the previous studies, we expect significant differences in achievements between low and high economic status students.

ONLINE ASSESSMENT OF SCIENTIFIC, INDUCTIVE REASONING SKILLS AND MOTIVATION TO LEARN SCIENCE

SAMPLES AND PROCEDURES

The sample of sub-study 1 was drawn from the fifth and seventh graders (N=616; 268 boys; 348 girls; age M=12.40, SD=1.19) from five different schools in Oshakati and Ongwediva towns. The schools were selected because of the availability of ICT infrastructure, but it was discovered upon arrival at the schools that the ICT facilities were malfunctioning.

The data collection was done through the eDia (Electronic Diagnostic Assessment) system (Molnar, 2015; Molnar & Csapó, 2013). The students had to be transported to the University of Namibia ICT rooms. Descriptive statistical analyses were used, together with classical statistic procedures such IRT (item response theory). Linear regression analysis was also applied.

RESULTS AND DISCUSSIONS

The reliability index of the scientific reasoning skills test was acceptable (Cronbach-alpha=.74) for the whole sample, however at the subscales level it yielded very low-reliability indices such that it may not be able to be interpreted. The 5th grade reliability index was low (Cronbach alpha= .69) compared to the 7th grade (Cronbach alpha= .72). For the science motivation questionnaire, the reliability index (Cronbach alpha= .91) was quite high for the whole sample and per grade. Furthermore, the scientific reasoning skills test was moderately hard for the students (M=37.83%; SD=13.34%), though they performed quite well in the proportional and correlational subtests compared to the rest of the subtests (above M=40%). The reliability of the domain general thinking skills was good compared to the scientific reasoning. The analyses showed (Cronbach alpha = .80) and all four subscales showed acceptable internal consistency of (Cronbach alpha >. 70). The students' performance in both skills tests was below 40%.

Rasch analysis was also carried out in order to gain a more detailed picture of the test. The EAP/PV reliability was about .70, which is acceptable. Further investigation showed that a few items were not suitable for differentiating students at low skill levels. This means that the test was a bit hard for the students. The 7th grade performed significantly better than the 5th grade ($t(616) = 7.87$, $p < .01$). Few students were at the lower end of the distribution, and few high ability items at the top.

Nonetheless, positive correlations were found between the subtests and the scientific reasoning achievement. Strong positive correlations were observed between conservation of mass and volume and proportional reasoning subscales with the main scale (scientific reasoning). However, the correlational reasoning scale showed a weak correlation with the main scale. The distribution of the achievements revealed that none of the students achieved above 80% and only three students scored below 10%. Furthermore, the correlation between scientific reasoning and domain general inductive reasoning was stronger in the 7th grade ($r = .31$).

ASSESSING STUDENTS' SCIENTIFIC INQUIRY SKILLS ABILITIES USING AN ONLINE INSTRUMENT

SAMPLE AND PROCEDURES

The sample for sub-study 2 was drawn from a secondary school in the Omusati region, in the north of Namibia. The school offers grades 8 to 12 and accommodates students from nearby villages and around the whole northern part of Namibia. It is a multicultural (grammar) school. The participants were the 9th and 11th graders (N=118, 44 boys and 74 girls; age mean=16.42, and standard deviation=1.25). The school has a hostel (dormitory), and all the learners were accommodated in the school hostel at the time of the study.

The online data collection was carried out through the eDia platform in January 2017, via internet in the school's information communication technology (ICT) room. Each participant was assigned a number (participant's ID) to log into the eDia system. Students entered their responses through the keyboard, choosing the right answer by clicking or by dragging and dropping figures on the screen with a mouse.

RESULTS AND DISCUSSIONS

The reliability of the whole scale was high (Cronbach alpha = .94), and two subscales emerged with a better reliability Cronbach alpha of .94 and .85 respectively. However, the reliability of the remaining five subscales needs improvement as this fluctuated by a (Cronbach alpha score of ≥.70). The range of scores on the pilot test was from 11% to 100%. The test mean score was (66.70%) with a standard deviation of 15.55% and a standard error of measurement of 1.43%. The test's Cronbach alpha almost matched that of the inquiry literacy test by Wenning (2007), when it was first piloted in the USA (Cronbach alpha =.88). The test was self-developed, though based on the framework by Korom et al. (2012).

Table 2 presents the correlation matrix showing bivariate relationships between the variables and the whole scale. Moderately strong positive correlations were found among the subscales. Strong correlations between the whole test and subscales were found. However, low correlations were found between the following sub scales i.e., experimental plans and hypothesis formulation; making conclusions and hypothesis formulation; than the rest the rest of the scales.

Table 2 Correlations coefficients between 7 sub scales and the whole test

	1	2	3	4	5	6	7
1. Whole scale (SI)	1						
2. Data handling	.75	1					
3. Identify variables	.78	.61	1				
4. RQs formulation	.82	.50	.65	1			
5. Hypothesis formulation	.67	.50	.45	.52	1		
6. Planning of variables	.90	.58	.61	.70	.58	1	
7. Experimental plans	.80	.54	.50	.58	.43	.73	1
8. Making conclusions	.80	.58	.51	.60	.42	64	60

Notes: all correlations are significant, * $p < .05$, ** $p < .01$

The Rasch analyses revealed that the distribution of the test was quite good. The difficulty level of the test fitted the students' ability. Students of different ability were discriminated fairly by the test. This shows that the test is good for measuring the students in the region. No significant differences were found between grades and genders concerning performances in SI. One-way ANOVA results showed that students whose mothers did not finish primary school performed significantly better than those whose mothers had reached the level of secondary or higher education ($F = 4.19$, $p < .05$). This is contrary to what is found in the literature.

THE RELATIONSHIPS AND ASSESSMENT OF SCIENTIFIC REASONING, SCIENTIFIC INQUIRY AND INDUCTIVE REASONING

SAMPLES AND PROCEDURES

The samples were chosen from larger secondary schools (8th to 12th grade schools). The students who participated in the study were in the 10th & 12th grades. The youngest age when the tests can be used is the 7th grade of primary school; below this age, reading difficulties are anticipated (Csapó, 1997). The oldest age group that can be tested within the educational system are 11th graders and beyond.

The data collection was done through the paper and pencil method. Students were given around two hours to complete each test. The researchers with the help of school teachers scored the test. The papers were marked and results entered into an SPSS file for quantitative analysis.

RESULTS AND DISCUSSIONS

Analyzing the test using descriptive statistics, the results showed that the internal consistencies of the tests were quite good (Cronbach alpha>.80 in all tests). Analysis of the whole sample showed a better performance in all three tests than in the previous sub-studies, where online methods were used. This finding is in line with the literature; for example, Bunderson, Inouye, and Olsen (1989) state: "the scores on tests administered on paper were more often higher than on computer-administered tests ... [though] the score differences were generally quite small" (p.378). In a meta-analysis of well-designed computer versus paper-based cognitive ability tests, Mead and Drasgow (1993) also found that on average, paper-based test scores were very slightly greater than computer-based test scores.

The relationships among SR, IR & SI were also examined. The SEM model of three domains that assess domain general reasoning skills (IR) and domain specific reasoning skills, i.e. SR and SI, showed a good model fit (see Figure 2). IR, SR and SI have been specified as latent factors. The results revealed that four dimensions, figure series (FS), figure analogy (FA), number series (NS) and number analogy (NA) explained the inductive reasoning (IR) performance ($.57 < \beta < .79$). Five dimensions explained the performance in scientific reasoning (SR): conservation of mass and volume (CON), proportional reasoning (PROP), control of variables (COV), probabilistic reasoning (PROB) and hypothetical deductive reasoning (HYP) ($.60 < \beta < .82$). Performance in the third domain of scientific inquiry skills (SI) was explained by seven dimensions: data analysis (DA), variable identification (VI), research questions identification (RQs), hypothesis identification (HF), variable planning (VP), planning experiments, (PE) and making conclusions (MC) ($.49 < \beta < .64$). Significant correlation was found between SI and IR ($r=.35$, $p<.01$).

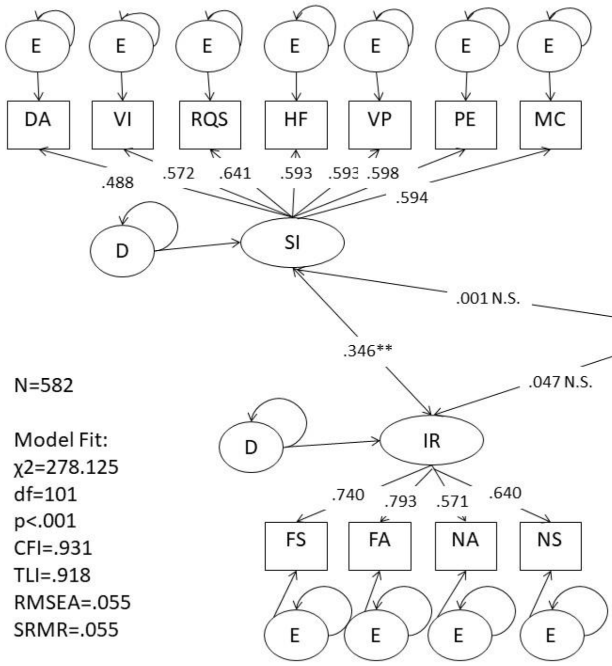

Figure 2. A model of SR, SI and IR with dimensions explaining each domain.

According to the literature, students from good socio-economic backgrounds tend to perform better in school subjects than those from poor socio-economic backgrounds (Keller, Neumann, & Fischer, 2017). ANOVA showed that there was no significant influence from the parents' level of education on SR and SI performances. However, significant difference was found between the groups *(mothers who did not finish primary school and mothers who finished only primary school)*, in the case of IR (p<. 05). Students whose mothers did not finish primary school performed better than those whose mothers studied up to secondary school and beyond.

THE RELATIONSHIPS BETWEEN 8TH GRADERS' SECOND LANGUAGE READINGS SKILLS, INDUCTIVE REASONING, SCIENTIFIC REASONING AND SOCIO-ECONOMIC STATUS

SAMPLE AND PROCEDURES

As in sub-studies 1, 2 and 3, the sample for sub-study 4 was drawn from northern Namibia public schools, representing students in four statistical regions of the country and schools located in settlements of different sizes. The sample comprised 250, 8th grade students (36% boys) and (64% girls), age mean (M=15.10; SD=.580). A reading comprehension test and a scientific and inductive reasoning test were used. The data collection method was also paper and pencil.

RESULTS AND DISCUSSIONS

Basic psychometric analyses revealed that the tests were reliable and an apt measuring instrument for this age cohort. Except for SR (Cronbach alpha=.74) the other two tests all yielded a Cronbach alpha of above .80. No significant gender difference was found in students' performance in all three tests. The comparison of students' RC, IR and SR ability scores using mean, standard deviation, and independent t-test showed that there were no significant differences in scores between genders ($p<.05$). These results resonate well with previous research that found no significant difference in the scientific and inductive reasoning ability between genders. Paired sample t-test results indicate that the sample performed significantly better in IR than in SR ($t_{(249)} = 6.09$, $p < .05$). The reasons for this may emanate from the fact that in the IR test not so much reading is involved, and only thinking and reasoning are required in order to derive the rules and get the answer, while with SR, a bit of reading and comprehension of some technical words is required, in addition to the reasoning part.

Additionally, the results indicate that the sample performed significantly better in the RC than in both IR and SR ($t=5.05$; $t=11.24$, $p<.05$ respectively). The above average performance in the RC test indicates that the sample did not have difficulty in understanding and following instructions in English as a second language. This finding reinforced the results obtained in sub-study 3 and resonates well with what is found in the international literature (cf. Nikolov & Csapó, 2018).

CONCLUSIONS

Through four complementary sub-studies, this study offers an assessment perspective on some of the skills that students need to acquire today. Since Namibia has not yet taken part in international assessment programs, (e.g. PISA, TIMSS), the aim was to ascertain students' abilities in scientific reasoning, scientific inquiry and inductive reasoning skills. As the increasing gap between girls and boys has been of some concern internationally, this study also focused on the possibility of gender differences. Another topic of special interest was whether students' fluency in the medium of instruction influenced their achievement in SR, SI and IR.

Sub-study one was one of the first attempts to carry out an online assessment in Namibia at primary school level. Although the literature suggested that technology-based assessment may provide schools and teachers with a user-friendly instrument for monitoring the development of students' thinking skills (Pásztor et al., 2015), the same may not be true for Namibia, as many public schools do not have suitably functional ICT infrastructure for online assessments to be an everyday school practice.

Furthermore, psychometric analyses of the reasoning skills tests revealed that the SR test must be revised in order to carry out more reliable assessment in this age cohort. The IR test reliability indices were high, although the students' performance was poor. The Rasch analyses revealed that the tests were hard for the primary school students. The task analysis revealed that students were not able to solve simple tasks and or find the rules in the case of the IR test. Their performance of these tasks highlights the need to explicitly expose students to and train them in different types of reasoning and thinking skills during the teaching and learning of STEM subjects. The results also highlight that both SR and IR develop with age, as older cohorts performed significantly better than their younger counterparts. Although there was weak correlation between scientific reasoning skills and the motivation to learn science, the results suggest that students are still keen to learn science in a general sense.

With regard to sub-study 2, the online assessment instrument for scientific inquiry skills proved to be reliable for the whole test. However, there were three subscales that yielded a low reliability index (Cronbach alpha of < .65), indicating a need for improvement if the subscales are reliably to measure the same constructs. The first step to be taken is to make the instrument suitable for everyday school practice and for large-scale assessments in Namibia. The findings indicate that online assessment may provide teachers with an easy-to-use instrument for monitoring the development of students' scientific inquiry skills and reasoning skills, and may contribute to the development of effective teaching and learning methods.

The results of sub-study 3, where the paper and pencil assessment method was used, proved to be very reliable as all the three tests yielded very high internal consistency of Cronbach alpha > .87 in both age groups. With regard to scientific reasoning, the original Lawson Classroom Test for scientific reasoning (LCTSR) was used (Lawson, 1978, 2000). The reliability results matched what was found by Lawson: (Cronbach alpha = .78, .77 & .80), respectively, on the same age group in the USA. The findings also suggest that hypothesis testing is a more advanced ability that students start to develop in their high school years. It is also the most rapidly changing (discriminating) ability among all the skill dimensions of SR. Regarding this skill dimension, the ceiling of the Lawson's test questions is a little over 80%. Two potential causes of this low ceiling were observed in the research. One is the length of the reading, which often causes students to lose track of the relevant experimental structure and variables and misinterpret the question. The other is related to the contextual elements of the questions. Students often tried to use their prior knowledge about red cells to answer the question pair 21-22, rather than using reasoning. Question pair 23-24 makes mention of a plastic bag that is semi-permeable. This is against common sense, as most plastic bags encountered in real life are waterproof. Therefore, some students thought the designs were implausible, and this prevented them from reasoning further.

Sub-study 4 was carried out to ascertain whether students' performances in the previous sub-studies was affected by the language issue. Sub-study 4's findings revealed that students performed above average in all three tests, a good result given that these cognitive and psychological tests were the first of their kind in Namibia. The results indicate that students achieved significantly better in the reading comprehension test than in the inductive and scientific reasoning tests, which means that students have no difficulty in following instructions in English. This finding is in line with one of the international assessments conducted in Hungary in 2015, where students' achievements afforded them the opportunity to study in the more desirable English language, associated with high cultural capital (Nikolov & Csapó, 2016).

The literature also suggests that parents' education predicts their children's educational outcomes, alongside other family characteristics such as the family income, parents' occupations, and residence location. Interestingly, the results from all the four sub-studies indicate the opposite. Students whose mothers did not finish primary education performed significantly better than those whose mothers had attained a secondary or higher level of education. One explanation for this is that, in Namibia, children from a low-income background tend to work harder than those from an affluent one, because of their desire to break out of poverty and live a better life. On the other hand, children of better educated parents do not seem to see the need to work so hard, since most of their needs are already met. This is merely a hypothesis that needs to be tested by further scientific research.

The study recommends a nation-wide study on how many public school have access to ICT. This is because the literature suggested that technology-based assessment may provide schools and teachers with a user-friendly instrument for monitoring the development of students' thinking and reasoning skills. Explicit training on SR, SI and IR is recommended for both teachers and students in order to improve their reasoning abilities and thus their performance in the STEM subjects. Results from sub-studies one and two suggest that some questions might have been left unanswered, like the relationship between the content of the tests and the Namibian 4-7 science curriculum. Further research is required to investigate this. A longitudinal study is also suggested, to find out whether it is true for all Namibian students that students from low SES performed significantly better than students from high SES.

REFERENCES

Adey, P. S., & Shayer, M. (1994). *Really raising standards*. London: Routledge.

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.

Bao, L., Cai, T., Koenig, K., & Fang, K. (2009). *Learning and scientific reasoning*. Washington, DC: AAAS.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., & Rumble, M. (2012). Defining 21st century skills. In P. Griffin, B. McGaw & E. Care (Eds.), *Assessment and teaching of the 21st century skills* (pp, 17-66). New York: Springer. Retrieved from http://atc21s.org/wp-content/uploads/2017/03/25-Defining-21st-Century-Skills.pdf

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (pp. 367-407). Washington, DC: American Council on Education.

Bybee, R. W. (2009). *The BSCS 5E instructional model and 21st century skills*. National Academies Board on Science Education, Washington, DC: Retrieved March 25, 2017, from https://www7.nationalacademies.org/bose/1Bybee_21st%20Century_Paper.pdf

Coletta, V. P., & Phillips, J. A. (2005). Interpreting FCI scores, normalized gain, reinstruction scores, scientific reasoning ability. *American Journal of Physics*, 1172-1179.

Csapó, B. (1997). Development of inductive reasoning: Cross-sectional measurements in an educational context. *International Journal of Behavioral Development*, *20*(4), 609-625.

Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology*, *106*(3), 639.

Fenichel, M., & Schweingruber, H. (2010). *Surrounded by science: Learning science in informal environments.* Washington, D. C.: National Press.

Han, J. B. S. (2013). *Scientific reasoning: Research, development, and assessment* Unpublished doctoral dissertation, The Ohio State University, Columbus, USA.

Harlen, W. (2013). *Assessment and inquiry-based science education: Issues in policy and practice*. Trieste: TWAS-Strada Costiera.

Hawkins, J., & Pea, R. D. (1987). Tools for bridging every day and scientific thinking. *Journal of Research in Science Teaching*, *24*(4), 291-301.

Inductive Reasoning: A training approach. Downloaded April 10, 2018, fromhttps://www.researchgate.net/publication/242148990_Inductive_Reasoning_A_Training_Approach

Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of Physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science teaching*, *54*(5), 586–614. doi: 10.1002/tea.21378

Klahr, D. (2000) *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.

Klauer, K. J. (1989). Teaching for analogical transfer as a means of improving problem-solving, thinking and learning. *Instructional Science, 18*(3), 179-192.

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. (1996). Teaching inductive reasoning: Some theory and three experimental studies. *Learning and Instruction*, *6*(1), 37-57.

Klauer, K. J. (1999). Fostering higher order reasoning skills: The case of inductive reasoning. In J. H. M. Hamers, J. E. H. van Luit & B. Csapó (Eds.), *Teaching and learning thinking skills* (pp. 131-156). Lisse: Swets & Zeitlinger.

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. J., & Phye, G. D. (2008). Inductive reasoning. A training approaches. *Review of Educational Research*, *78*, 85-123.

Klauer, K. J., Willmes, K., & Phye, G. D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology*, *27*(1), 1-25.

Korom, E., Németh, M., Pásztor, A., & Csapó, B. (2017). Relationship between scientific and inductive reasoning in grades 5 and 7. *17thBiennial Conference of the European Association for Research on Learning and Instruction (EARLI)* (pp. 128-129). Tampere, Finland, August 29-September 2, 2017.

Korom, E., Nagy, L., Nemeth, M. B., Radnoti, K., Farkas, M. A., Markocsi, I. R., & Wagner, E. (2012). Detailed frameworks for diagnostic assessment of science. In B. Csapó & G. Szabo, *Frameworks for diagnostic assessment of science* (pp. 175-306). Budapest: Nemezti Tankonyvkiado.

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 to large scale assessment* (pp. 13-23). Brussels: European Communities.

Kuhn, D. (2010). What is scientific thinking and how does it develop? In U. Goswami, *Handbook of childhood cognitive development* (pp. 497-523). Oxford: Wiley Blackwell.

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, *15*(1), 11-24.

Lawson, A. E. (2000). Development and validation of the classroom test of formal reasoning. *Journal of Research in Science Teaching*, *15*(1), 11-24.

Lazonder, A. W., & Kemp, E. (2012). Bit by bit or all at once? Splitting up the inquiry task to promote children's scientific reasoning. *Learning and Instruction*, *22*(6), 458-464.

Mayer, D., Sodian, B., Koerber, S., & Schwipert, K. (2014). Scientific reasoning in elementary school children: Assessment and relation with cognitive abilities. *Learning and Instruction*, *29*, 43-55.

Mayer, J. (2007). Inquiry as scientific problem solving. In D. Kruger & H. Vogt (Eds.), *Theories in biology didactic* (pp. 177-186). Heidelberg: Springer.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*, 449-458.

Ministry of Education. (2007). *Education and training sector and improvement program*. Windhoek: Government Press.

Nikolov, M., & Csapó, B. (2010). The relationship between reading skills in early English as a foreign language and Hungarian as a first language. *International Journal of Bilingualism, 14*, 315-329.

Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, *10*, 265-279.

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.

Vainikainen, M. P. (2014). *Finnish primary school pupils' performance in learning to learn assessments: A longitudinal perspective on educational equity*. Unpublished doctoral thesis, University of Helsinki, Helsinki Finland.

Wenning, C. J. (2007). Assessing inquiry skills as a component of scientific literacy. *Journal of Physics Teacher Education Online*, *4*(2), 21-24.

Yeh, Y., Jen, T., & Hsu, Y. (2012). Major strands in scientific inquiry through cluster analysis of research abstracts. *International Journal of Science Education*, *34*, 2811-2842.

Zimmerman, C. (2005). *The development of scientific reasoning skills: What psychologists contribute to an understanding of elementary science learning?* National Research Council.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*, 172-223.

PUBLICATIONS RELATED TO THE DISSERTATION

Julius, H. L. & Kambeyo, L. (2016). Investigating the motivating factors that influenced the University of Namibia first year undergraduate student teachers to choose teaching as a career: A case study. In H. Miranda (Ed.), *Proceedings of the 5th UNAM Annual Educational Conference* (pp. 167-168, 22-24 August 2016, Ongwediva, Namibia.

Kambeyo, L., & Wu, H. (2018). Online assessment of students' inductive reasoning skills abilities in Oshana Region, Namibia. *International Journal of Educational Sciences, 20*(1), 137-147.

Kambeyo, L., & Wu, H. (2017). Nemzetközi tapasztalatok a tanulói kompetenciák mérésére. In G. Hunyady, B. Csapó, G. Pusztay & J. Szivák (szerk.), *Az oktatás korproblémái*. (pp 179-189) ELTE Eötvös Kiadó, Budapest.

Kambeyo, L., & Csapó, B. (2017). The feasibility of online assessment of students' inductive reasoning skills abilities in Namibia. *Proceedings of the 9th EDULEARN17 Conference* 3rd-5th July 2017, Barcelona, Spain.

Kambeyo, L., & Csapó, B. (2018). Scientific reasoning skills: A theoretical background on science education. *Reform Forum Journal for Educational Reform in Namibia*, *26*(1), 29-39.

Kambeyo, L. (2016). Scientific reasoning skills: A theoretical background on science education. In H. Miranda (Ed.), *Proceedings of the 5th UNAM Annual Educational Conference* (pp. 192-196), 22-24 August 2016, Ongwediva, Namibia.

Kambeyo, L. (2017) Assessing students 'abilities of scientific inquiry skills. *6th Interdiszciplináris Doktorandusz Konferencia (IDK2017*) (6th Interdisciplinary Doctoral Conference), Pecs, Hungary, 19-21 May, 2017.

Kambeyo, L. (2017). Online assessment of scientific reasoning skills: The case of Namibia. *15th Conference on Educational Assessment (CEA*), Szeged, Hungary, 6-8 April, 2017.

Kambeyo, L. (2017). The possibilities of assessing students' Scientific Inquiry Skills Abilities Using an Online Instrument: A Small-Scale Study in the Omusati Region, Namibia. *European Journal of Educational Sciences, 4*(2), 1-21.

Kambeyo, L. (2018). The assessment of English reading comprehension and its predictive power on inductive reasoning. *16th Conference on Educational Assessment (CEA),* 26-28 April 2018, Szeged, Hungary.

Kambeyo, L. (2018). The assessment of scientific reasoning, inductive reasoning and scientific inquiry using paper and pencil in Namibia. *Proceedings of 10th EDULEARN18 Conference, 2nd-4th, July 2018,* Palma de Mallorca Spain. Paper Accepted

Kambeyo, L. (2018). The assessment of scientific reasoning, inductive reasoning and scientific inquiry using paper and pencil in Namibia. *16th Conference on Educational Assessment (CEA),* 26-28 April, 2018, Szeged, Hungary.

Kambeyo, L., Pásztor, A., Korom, E., & Németh, B. M. (2018). Online assessment of thinking skills in science context and motivation to learn science among the grade 7 students in Oshana region, Namibia. *26th Annual Conference of the Southern African Association for Research in Mathematics, Science and Technology Education (SAARMSTE)* (p. 129), Gaborone, Botswana, 16-19 January 2018.

Kambeyo, L., Pásztor, A., Korom, E., Németh, B. M., & Csapó, B. (2017). Online assessment of scientific reasoning and motivation to learn science: A pilot study in Namibia. *17th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI)* (p. 129), Tampere, Finland, August 29- September 2, 2017.

Orosz, Z. G., B. Németh, M., Kambeyo, L., Korom, E. (2017). Assessing inquiry skills in Hungary and Namibia. In J. Kerülő, T. Jenei, I. Gyarmati (Eds.), *ONK 2017. XVII. National Educational Conference* (p. 532). Nyíregyháza, Hungary, 9-11 November, 2017.

Wu, H., Kambeyo, L., & Molnár, G. (2017). The feasibility and reliability of computer-based assessment of inductive reasoning in Namibia and China: A comparison study.*15th Conference on Educational Assessment (CEA)*, 6-7April 2017, Szeged, Hungary.