INFLUENCE OF SCIENCE TEACHING AND ASSESSMENT MODALITIES ON STUDENTS' ELEMENTARY SCIENCE PERFORMANCE

Evrim GENÇ KUMTEPE*

FEN ÖĞRETİMİ VE ÖLÇME YAKLAŞIMLARININ ÖĞRENCİLERİN İLKÖĞRETİM FEN BAŞARISI ÜZERİNE ETKİSİ

ÖZET

Bu çalışmanın amacı, fen öğretim ve ölçme uygulamalarının ilköğretim düzeyindeki öğrencilerin fen başarısına etkisini incelemektir. Çalışmada, Okul Öncesi Uzun Dönem Araştırması-Anaokulu Sınıfı 1998-99 verileri (ECLS-K) kullanılmıştır. İlköğretim dönemi çocukların verilerini içeren bu çalışmaya ilişkin veri seti 2004 yılında yayımlanmıştır. Öğrenci ve öğretmen düzeyi değişkenlerinden cinsiyet, sınıf ve fen öğretimi ve ölçme tekniklerinin (üst düzey düşünme becerileri, fen etkinliklerine ayrılan zaman ve test tabanlı ölçme uygulamaları gibi) öğrenci fen bilgisi başarısına etkilerini araştırmak amacıyla çoklu regresyon modeli uygulanmıştır. Önerilen regresyon modeli, öğrencilerin fen bilgisi başarısındaki varyasyonun yaklaşık %11'ini istatistiksel olarak önemli düzeyde açıklamıştır. Ayrıca, modelin açıklanmasında, tüm bağımsız değişkenlerin etkisi istatistiksel olarak anlamlı bulunmuştur. Özellikle erkek öğrencilerin ve üçüncü sınıfların fen başarısı kız öğrencilerden ve ikinci sınıflardan önemli düzeyde yüksektir. Ayrıca, öğretmenler alternatif ölçme tekniklerini kullanmaya ve analiz, sentez ve değerlendirme gibi üst-düzey düşünme becerilerine odaklandıklarında, öğrenciler fen dersinden daha fazla yararlanmıştır.
Anahtar Sözcükler: Fen öğretimi, fen başarısı, üst düzey düşünme becerileri, ECLS-K

ABSTRACT

The aim of the study is to investigate the influence of science instruction and assessment practices on elementary level students' science success utilizing the data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K). The specific data set of the study regarding elementary school age children's variables was released in 2004. A multiple regression modeling technique was employed to explore the effects of students and teacher level variables, including gender, grade level, science teaching and assessment techniques (such as emphasizing higher-order thinking skills, the time allocated for science-related activities, and test-based assessment practices) on students' science achievement. Regression analysis of the proposed model revealed that the model significantly explained about 11% of the variance on students' science scale scores. Furthermore, all predictors significantly contributed and estimated students' science performance in the model. Specifically, the boys and the third graders had significantly higher mean science scores than the girls and the second graders. Moreover, when teachers tend to implement alternative assessment methods and focus more on higher-order skills such as analysis, synthesis, and evaluation in science, students benefited more in science.
Keywords: Science teaching, science achievement, higher-order thinking skills, ECLS-K

* Assoc. Prof. Dr. Distance Education, Open Education Faculty, Anadolu University. Office No: 607, Eskişehir, Turkey. egkumtepe@anadolu.edu.tr

1. INTRODUCTION

The National Science Education Standards [NSES] and the National Association for the Education of Young Children [NAEYC] affirm that accessible and high-quality science education for 3- to 10-year-old children is the central underpinning for future science learning (NAEYC, 2001; NRC, 1996). These documents highly emphasize that teaching and assessment should be redesigned to promote student understanding through scientific inquiry. They also point to fact that effective science teaching requires understanding of what students know and need to learn and then challenging and supporting them to learn it well. In the same vein, assessment should support children's science learning and provide useful information to various stakeholders, including teachers, students, and community.

Teaching and assessment may be practiced in many different ways, depending on the perspective one takes. Every teacher develops a particular way of progress about the demanding task of teaching and assessment. One introduces a topic by raising questions to capture students' existing knowledge or misconception(s) about a specific topic whereas another teacher only gives a lecture and then a quiz to evaluate students' understanding. One puts more emphasis on knowing and learning basic skills in science whereas other implements blended teaching and assessment method not only focusing on fact of knowledge but also on understanding students' higher-order learning skills in science such as problem solving and critical thinking. *Understanding* is a complicated term "stand-alone" and *self-understanding* is, therefore, a complex process that is not easily measured and is directly related to the nature of human mind. The question of *How do people think and learn?* has been viewed as one of the most influential questions of science since the 20th century (Pellegrino, Chudowsky, & Glaser, 2001).

Researchers from a variety of disciplines- such as anthropology, linguistics, cognitive science, developmental psychology, and biology- have been studying the nature of human mind and developing theories to expand our understanding of such matters: how mind works, how a child develops conceptual understanding, and how people think, know and learn, in particular. Concerning science teaching and assessment, all advances in the theories of knowing and learning, now, assist us to realize diverse perspectives and understandings of human thinking with differing implications for what should be taught and how should be assessed (Greeno, Pearson, & Schoenfeld, 1996; Pellegrino et al., 2001). According to Greeno et al. (1996), the four perspectives on the nature of the human mind have considerable influence in the history of research and theory: *Behaviorist, Differential, Cognitive, and Situative Perspectives.* The first two perspectives reflect the most widely used traditional teaching and assessment ways. Teachers having these perspectives generally employ teacher- or textbook-based tests to measure student science comprehension. They also put more emphasis on remembering or memorizing scientific facts instead of assessing higher-order skills in science. Alternatively, cognitive and situative learning perspectives reveal the importance of the development of knowledge and sociocultural dimension of learning (Pellegrino, et al., 2001). As described by Webb and Mason (2003), these teaching and learning perspectives are being recognized as alternative views of science teaching practices. Science teaching and assessment of learning should be based on higher-order thinking skills so that students can cope with real-life related problems

(Winking, 1997). These skills in science are seen as the key for learning in that students make certain learning possible and make acts of carrying out certain tasks in science. In this respect, a framework that is still regarded as being functional was devised by Bloom (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). Concerning his taxonomy of knowledge, summarized in Figure 1, moving from knowledge acquisition (usually considered as a lower order thinking skill) up to analysis, synthesis, and evaluation (often considered as higher-order thinking skills) phases, knowledge progresses from simpler to more complex forms. During this process, we want to foster students' abilities not only to understand, but also to engage in inquiry and to apply scientific thinking to problems in life. Probably an unquestionable issue, herein, is the fact that science teaching should be based on experimentation (hands-on, minds-on), higher-order process skills, and new discoveries.

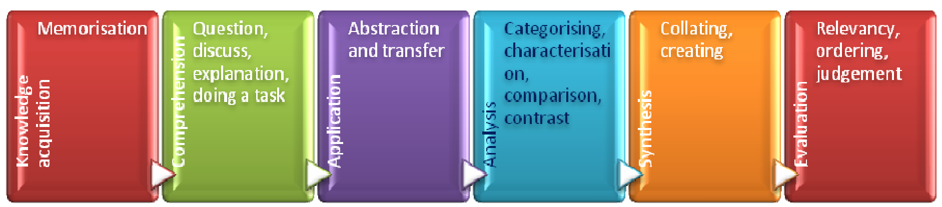

Figure 1. Bloom's taxonomy of knowledge and associated scientific process skills

The learning perspectives in terms of traditional and alternative forms of teaching and assessments are not seen as independent viewpoints in the NSES and the NAEYC because both assist educators to find out what students know and can achieve. However, an increasing number of research have been recommending the use of cognitive and situative views of knowing, herein referring to alternative teaching techniques, in an attempt to expand the knowledge on students' science learning (NRC, 1999; NAEYC, 2001; Pellegrino, et al., 2001; Webb & Mason, 2003). This study therefore is an attempt to examine the influence of alternative and innovative science instruction and assessment techniques on elementary students' science achievement utilizing the data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K). Specifically, the study addressed the following research questions:

1. What is the association between science teaching factors and students' science achievement at elementary level?
    a. What is the association between the student level factors including gender and grade level and elementary students' science achievement?
    b. What is the association between the time allocated for science related activities and elementary students' science achievement?
    c. What is the association between the degree of emphasis on teaching higher order skills and elementary students' science achievement?
    d. What is the association between the frequency of using science related instructional activities and elementary students' science achievement?
    e. What is the association between the frequency of using test-based assessments and elementary students' science achievement?

2. METHODOLOGY

2.1. Data source and sample

The data of the study is derived from the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), conducted by National Center for Educational Statistics (NCES). ECLS-K is a "multisource, multimethod study that focuses on children's early school experiences beginning with kindergarten" (US Department of Education, 2000b, p, 1-1). The ECLS-K is a longitudinal study that gathers data from a nationally representative sample of nearly 22,000 children from kindergarten through the eight grade. The children in ECLS-K came from both public and private schools and diverse socioeconomic and racial/ethnic backgrounds. Parents, teachers, and school administrators are also involved in the study. ECLS-K is not a standardized or mandatory state or national test to be performed. The idea for the study is to produce a comprehensive and reliable data source that can provide researchers, teachers, parents, and policy makers to better portray and to understand children's progress from kindergarten through middle school years. The data set of the current study was released in 2004 for researchers' use.

The working sample was generated from child data if a child had complete records on study variables in 2nd and 3rd grade levels. The last working sample included 9,463 elementary students and 3,556 teachers. Since ECLS-K is not a simple random sample, which means not all schools, teachers, and children had an equal probability of selection; an appropriate student weight was initially normalized and then adjusted for the analysis. Using SPSS 13.0 (Statistical Program for the Social Sciences), the standard errors were corrected with average root design effect (DEFT) to calculate standard errors, assuming the data were collected with a simple random sample (SRS). In the SPSS, the standard errors were corrected using DEFT. The standard error of an estimate under the actual sample design was approximated with the following formula;

$$SE_{DESIGN} = \sqrt{DEFF \times Var_{SRS}} = DEFT \times SE_{SRS}$$

After using the appropriate study weight for children data, the sample was nationally representative of 1,205,271 children.

2.2. Study variables

The science assessment domain in ECLS-K study includes items that measure knowledge and skills in the natural sciences. During assessment, it is reported that equal emphasis was placed on earth and space science, physical science, and life science. The variables in the current study included both student and teacher level characteristics. Variables and their operational definitions are listed in Table 1. Third graders' science achievement is the outcome of this study (dependent variables). The variables related to students' and teachers' characteristics were used as predictors (independent variables) in the model.

Table 1. Description of the variables used for the analysis

	Outcome Variable
Science Achievement	Students' IRT scale scores in science[†]
	Predictors
Student/ Class Characteristics	
Gender (X_1)	Dichotomously coded (Male=1, Female=2)
Grade Level (X_2)	Dichotomously coded (2nd grade=1, 3rd grade=2)
Science or Nature Area (X_3)	Reflecting an existing science or nature area (Dichotomously coded; No=1; Yes=2)
Teacher Characteristics	
How often science (X_4)	Indicating the time spent by students working on lessons or projects in science classrooms (Never=1 through Daily=5)
Higher-Order Scientific Process Skills (X_5)	Composite variable indicating the degree of emphasis on teaching higher order skills in science[‡]
Enriched Science Activities (X_6)	Composite variable indicating frequency of using various science related instructional activities[§]
Test-Based Assessment (X_7)	Composite variable reflecting frequency of using three different traditional assessments in science

2.3. Analysis

Descriptive statistics were performed to examine means, standard deviations, percentages, and frequency distributions of the variables in the data. A linear multiple regression analysis was employed to estimate the accounted variances of the students' science achievement by the selected variables, including student level and teacher level predictors. The inclusion of multiple predictors in a regression model allows the use of statistical control in estimating the unique effects of these predictors on student science achievement. Descriptive statistics for the outcome and the predictors in the regression model are tabulated in Table 2 below.

[†] Item Response Theory (IRT) based scale scores in science. IRT uses the pattern of right, wrong, and omitted responses to the items actually administered in a test and the difficulty, discriminating ability, and "guess-ability" of each item to place each child on a continuous ability scale. It is then possible to estimate the score the child would have achieved if all of the items in all of the test forms had been administered. By using the overall pattern of right and wrong responses to estimate ability, IRT can compensate for the possibility of a low ability student guessing several hard items correctly.

[‡] Including eight items on a 3 point Likert scale; emphasis on science facts, understand science concepts, develop science problem solving, learn science relevance, communicate science ideas, develop science lab skills, develop science interest, and develop data analysis skills (Little or no emphasis=1; Moderate emphasis=2; Heavy emphasis=3)

[§] Including nine items on a 4 point Likert scale; frequency of reading about science, discussing science news, science projects, science oral report, science written report, hands on science activity, talking about science results, using library science resources, freq using computer science (Never or hardly ever=1; Once or twice a month=2; Once or twice a week=3; Almost every day=4)

Table 2. Descriptive Statistics for Outcome and Predictors in the Regression Model Grouped by Grade Level ($n_{student}$=9,463; $n_{teacher}$=3,556)

Variables		Grade Level (X_2)			
		Second Grade ($n_{student}$ =737; $n_{teacher}$=579)		Third Grade ($n_{student}$ =8,726; $n_{teacher}$=2,977)	
		M	SD	M	SD
Science Achievement (Y)		26.16	8.7	35.39	9.3
		n	%	n	%
Student Gender (X_1)					
	Male	467	63.4	4,262	48.8
	Female	270	36.6	4,464	51.2
Science or Nature Area (X_3)					
	No	466	63.2	5,530	63.4
	Yes	271	36.8	3,196	36.6
		M	SD	M	SD
How often Science (X_4)		3.62	.81	3.73	.80
Higher Order Scientific Process Skills (X_5)		2.05	.43	2.18	.38
Enriched Science Activities (X_6)		2.06	.45	2.14	.42
Test-Based Assessment (X_7)		2.61	.62	2.68	.50

Note: M=Arithmetic mean of scores; SD=Standard Deviation; n:sample size (frequency distribution)

A total of 9,463 students' data were used in the analyses and approximately 50% of these students were male (n_{male}=4,729; n_{female}=4,734) students. The average second and third grade students' science scale score were found to be 26.16 and 35.39, respectively, based on item response theory scale that displays ability estimate parameters. Approximately 37% of the classrooms had a science or nature area in both grade levels. It was reported that second grade students frequently (three or four times a week) worked on lessons or projects in science classrooms. By the same token, the data showed that the third grade students were frequently engaged in various scientific activities (M_{third}= 3,73, considering mean scores on a 5 point scale). The mean of emphasis on teaching higher-order scientific process skills was on average based on a 3 point scale (M_{second}= 2.05; M_{third}= 2.18) for both grade levels. On the other hand, the data revealed that they did not frequently implement science related activities (M_{second}= 2.06; M_{third}= 2.14, considering mean scores on a 4 point scale). It is promising that teachers did not often use traditional mode of assessment, herein, named test-based assessments in science (M_{second}= 2.61; M_{third}= 2.68, considering mean scores on a 5 point scale).

3. RESULTS

3.1. Preliminary analyses

The proposed model was confirmed in a linear multiple regression analysis with student science achievement as the dependent measure and a total of seven student and teacher characteristics served as determinants. Initially, preliminary analyses, including an analysis of missing data, a case analysis to identify problematic individual observations, and an assessment of possible violations of the regression assumptions, were conducted. The subjects with missing data were excluded first. Five influential cases that had standardized residuals with an absolute value greater than 3 were identified as problematic and removed from the data. No significant outliers (Cook's distance <1) was observed. Regarding assumptions, all predictor variables and the outcome were measured at the interval level. All predictors in the model had some variation in value (see Table 2). There was no evidence of multicollinearity considering collinearity statistics (e.g., VIF and the tolerance values). Specifically, the largest VIF was estimated as 1.530 that is less than 10 and the average VIF was 1.17 which is not substantially greater than 1, therefore, there was no cause for concern (Bowerman & O'Connell, 1990; Myers, 1990). Moreover all the tolerance values were greater than .2, which indicates no collinearity problem in the data (Menard, 1995).

3.2. Model summary

Regression analysis revealed that the model significantly predicted students' science achievement, F (7, 9455) =171.84, p<.001. R^2 for the model was .113, and adjusted R^2 was .112, indicating that the proposed model including seven explanatory variables related to student characteristics and teachers' science teaching modalities explained 11.3 % of the variability on students' science achievement.

The list of interpretations for each of the predictors in the model

The following equation reflects estimates, β, obtained through the regression model given in Table 3.

Student Science Achievement= β_0+β_1 Gender + β_2 Grade Level + β_3 Science Area + β_4 How Often Science + β_5 Higher-order Skills+ β_6 Scientific Activities+ β_7 Test-based Assessment

$$\hat{Y} = 17.53 - 2.346\ X_1 + 9.4\ X_2 + .724\ X_3 + .732\ X_4 + 1.43\ X_5 + .921\ X_6 - 2.319\ X_7$$

According to Table 3, in terms of individual relationships between science achievement and each predictor, all predictors significantly predicted the outcome. Initially, the predicted average science score for male students were 2.46 units higher than that of female students, controlling for all other predictors in the model (β_1= -2.346 ; t=-12.645, p<0.001). The estimated science achievement for second graders were lower than that of third grade students by 9 points, controlling for all other predictors in the model (β_2= 9.4; t=27.033, p<0.001). Incorporating science or nature areas into science activities was positively related to student science achievement (β_3= .724 ; t=3.606, p<0.001). In other words, students benefitted more in science when they experienced science in nature areas. There is an expected .7 increase in science achievement when students frequently conducted science related projects (β_4= .732; t=4.943, p<0.001). When teachers put greater emphasis on higher-order scientific process skills in science (i.e., problem solving, data analysis, reasoning, interpreting), their students' scores on science were higher than that of other students whose teachers did not give importance on using these skills in lessons (β_5= 1.430, t=5.918, p<0.001).

Table 3. Results of Science Achievement Regression Model

	Unstandardized coefficients			95% Confidence Interval for	
	β	se	*t*	β	
Constant	17.530	1.007		15.556	19.505
Gender (X_1)	-2.346	.186	-12.645*	-2.710	-1.982
Grade Level (X_2)	9.400	.348	27.033*	8.719	10.082
Science or Nature Area (X_3)	.724	.201	3.606*	.330	1.117
How often science (X_4)	.732	.124	5.918*	.489	.974
Higher Order Scientific Process Skills (X_5)	1.430	.289	4.943*	.863	1.998
Science Enriched Instructional Activities (X_6)	.921	.270	3.415*	.392	1.449
Test-Based Assessment (X_7)	-2.319	.184	-12.622*	-2.679	-1.959

β=coefficients of the model; se= standard error; t=t statistics
** p <0.01

When teachers frequently use enriched science activities, their students achieved more in science (β_6= .921, t=-3.415, p<0.001). Finally, the results revealed that alternative assessment modes were superior to traditional techniques. More frequently using standardized tests, commercially produced test, and teacher-made tests/quizzes (β_7=-2.319, t=-12.622, p<0.001) resulted a decrease in science achievement.

4. DISCUSSION AND IMPLICATIONS

Research spanning several decades underline that effective teaching needs to supervise and perform a wide range of activities and skills which have been recognized as potential determinants of students' academic success. With respect to the teaching of science, scientific thinking skills, and assessment of learning, the results of this study replicate other studies in which innovative science instruction anchored in hands-on activities and performance based assessment was clearly superior to traditional lecture based science teaching in facilitating children's acquisition of scientific concepts (Chang & Barufaldi, 1999; Dimitrov, 1999; Dori, 2003; Genc, 2005; Martin, Mullis, Beaton, Gonzalez, Smith & Kelly, 1997; Stohr-Hunt, 1998; Wenglinsky, 2000).

Taken together, the hypothesized model described here is intended to develop a sense of what the effects of different science teaching and assessment modalities on elementary science achievement are. More specifically, it is believed that inspecting different examples of science teaching in a large scale data set (ECLS-K) provided insight into practical and innovative aspects of the teaching science such as improving students' observational skills and reasoning abilities, orchestrating meaningful discussions, and continually providing constructive feedback to learners.

Gender differences in science have been widely investigated in the science education research, comparing on a range of variables such as achievement, attitude, motivation, interest, and performance behaviors. In general, males are found to be more interested and successful in science as opposed to females (Osborne, Simon & Collins, 2003; Steinkamp & Maehr, 1983). The student level predictors of the present study also revealed consistent findings with previous research. More specifically, the results yielded that both boys and third graders had significantly higher mean science achievement than girls and second graders. By the same token, Martin, et al. (1997) reported that girls performed significantly less than boys at both the third and fourth grades internationally and in about half of the TIMSS (Trends in International Mathematics and Science Study) countries. Dimitrov (1999)'s study focused on gender differences as well but he further examined the differential effects of gender, ability, response formats, and learning outcomes on student science achievement. The results showed that boys did better than girls on the open-ended format in physical sciences at the high ability level, but no gender differences were observed in nature of science, earth and space sciences, and life sciences in particular. Furthermore, he reported no gender differences in science achievement for the low and medium ability students. Therefore, it is suggested that the current data may be reexamined in an attempt to determine if ability or performance levels in science is differentiated by gender.

The most important result of this study is the high degree of consistency between students' science achievement and the frequency of science experience, specifically hands-on activities. Previous studies also addressed this fact that hands-on activities, if regularly incorporated to classroom instruction, enhance cognitive learning (Echevarria, 2003; Freedman, 1997, Gerstner & Bogner, 2009; Stohr-Hunt, 1998). According to Stohr-Hunt (1998)'s study on the relationship between the amount of time students spent experiencing hands-on science and science achievement measured through the National Education Longitudinal Study of 1988, the significant differences existed across the hands-on frequency variable with respect to science achievement. Specifically, students who engaged in hands-on activities every day or once a week scored significantly higher on a standardized test of science achievement than students who engaged in hands-on activities once a month, less than once a month, or never (p.101). On the one hand, Gerstner and Bogner (2009) cautioned that "spontaneously implemented hands-on instructions did not evoke positive effect on achievement scores", although hands-on approaches motivate students' science learning (p.2). For this reason, it is suggested that teachers should evolve and incorporate activity-based learning settings in regular plans beforehand, to build up students' fascination for science.

The findings of the study also supported the common notion in research that conveying higher-order thinking skills lead to improved student science achievement (Chang & Barufaldi, 1999; Wenglinsky, 2000; Zohar & Dori, 2003). The results of Chang and Barufaldi (1999)'s study proved that the problem-solving-based instructional model did significantly improve students' science achievement comparing to the traditional-lecture approach. They also underlined that application of science concepts rather than rote learning or memorization should be emphasized in the science classroom to help students develop higher-level thinking skills (Chang & Barufaldi, 1999). In other words, teachers should pay more attention to engaging students in discussing, questioning, researching, problem-solving, and communicating their scientific ideas to enhance their learning in science.

On the subject of assessment practices, the data yielded similar results with previous studies as well. Specifically, when teachers tend to implement alternative assessment techniques rather than testing, students benefited more in science (Bol &

Strage, 1996; Dori, 2003; Genc, 2005). Meanwhile, previous research investigated the potential impact of high-stakes testing on teaching science in elementary to high schools (Genc, 2005; Hilliard, 2002; Pringle & Martin, 2005; Shepard, 2002). It is underscored that teachers constantly feel the pressure of high-stakes testing, even though they are willing to implement innovative assessment methods. For instance, the majority of science teachers in Florida claimed that the state-mandated accountability tests hindered their actual teaching (Genc, 2005). Furthermore, it is indicated that they are trying to align their classroom instruction with the science standards when aggressively searching for test preparatory materials (Pringle & Martin, 2005). Pragmatically, we must move beyond any type of tests including standardized tests to teacher-produced ones to more authentic measures of student understanding in science such as individual/group projects, self and peer assessments, oral reports, written assignments, and portfolios, etc.

Overall, it is believed that the results of this study, utilizing such a nationally representative data of "ECLS-K", would help policy makers plan more effective science programs so that young children's science learning and development would be enhanced. Still, such results could provide evidence-based guidance to teachers for implementing new instructional approaches to early science instruction. It is highly recommended that researchers continue to examine the data and the new releases of it and to generate reliable information for the future.

REFERENCES

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, Handbook I: Cognitive domain*. New York: Longmans Green.

Bol, L., & Strage, A. (1996). The contradiction between teachers' instructional goals and their assessment practices in high school biology courses. *Science Education, 80*, 145–163.

Bowerman, B. L., & O'Connell, R. T. (1990). *Linear statistical models: An applied approach* (2nd ed). Belmont, CA: Duxbury.

Chang C-Y., & Barufaldi J. P. (1999). The use of a problem-solving-based instructional model in initiating change in students' achievement and alternative frameworks. *International Journal of Science Education, 21*, 373-388.

Dimitrov, D. M. (1999). Gender differences in science achievement: Differential effect of ability, response format, and strands of learning outcomes. *School Science and Mathematics, 99*, 445–450

Dori, Y. J.(2003). From nationwide standardized testing to school-based alternative embedded assessment in Israel: Students' performance in the matriculation 2000 project. *Journal of Research in Science Teaching, 40*, 34-52.

Echevarria, M. (2003). Hands on science reform, science achievement, and the elusive goal of 'science for all' in a diverse elementary school district. *Journal of Women and Minorities in Science and Engineering*, *9*, 375-402.

Freedman, M. P. (1997). Relationship among laboratuary instruction, Attitude toward science, and achievement in science knowledge. *Journal of Research in Science Teaching, 34*, 343-357.

Gerstner, S. & Bogner, F. X. (2009). Cognitive achievement and motivation in hands-on and teacher-centred science classes: does an additional hands-on consolidation phase (concept mapping) optimise cognitive learning at work stations? *International Journal of Science Education*, Retreived July 28, 2009 from http://www.informaworld.com/smpp/content~db=all~content=a912679478

Genc, E., (2005). *Development and validation of an instrument to evaluate science teachers' assessment beliefs and practices. Unpublished doctoral dissertation*, Florida State University, Tallahassee, USA.

Greeno, J. G., Pearson, P. D., and Schoenfeld, A. H. (1996). **Implications for NAEP of research on learning and cognition. Report of a study commissioned by the national academy of education.** Panel on the NAEP Trial State Assessment, conducted by the Institute for Research on Learning. Stanford, CA: National Academy of Education.

Hilliard, A. G. (2000). Excellence in education versus high-stakes standardized testing. *Journal of Teacher Education, 51*, 293–304.

Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T .A., Kelly, D L. (1997). Science achievement in the primary school years: IEA's third international mathematics and science study (TIMMS). TIMSS International Study Center: Boston College, Chestnut Hill, MA, USA

Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage.

Myers, R. (1990). *Causal and modern regression with applications* (2nd ed). Boston, MA: Duxbury.

National Association for the Education of Young Children [NAEYC]. (2001). *NAEYC standards for early childhood professional preparation. Washington*, DC: Author.

National Science Teacher Association. (1996). Survey finds science teachers optimistic National Science Education Standards will foster change. Washington, D. C. Retrieved June 10, 2009 from http://www.nsta.org/survey1.

National Research Council [NRC]. (1999). *The assessment of science meets the science assessment*. The summary of Workshop. Washington, DC: National Academy Press,

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*, 1049-1079.

Pellegrino, J. W., Chudowsky, N., & Glaser, R.., Eds. (2001). *Knowing what students know: The science and design of educational assessment*. Washington D. C.: National Academy Press.

Pringle, R., & Martin, S. (2005). The potential impacts of upcoming high-stakes testing on the teaching of science in elementary classrooms. *Research in Science Education, 35*, 347-361(15).

Shepard, L. (2002). The hazards of high-stakes testing. *Issues in Science and Technology Online*. Retrieved on June 11, 2009, from http://www.issues.org/issues/19.2/shepard.htm

Steinkamp, W. M. Maehr, M. L. (1983). Affect, ability and science achievement: A quantitative synthesis of correlational research. *Review of Educational Research, 53*, 369-396.

Stohr-Hunt, P. M. (1998). An analysis of frequency of hands-on experience and science achievement. *Journal of Research in Science Teaching, 33*, 101 – 109.

Webb, N.L. & Mason, S. A. (2003). Taking stock of the national science education standards: the research for assessment and accountability. In Hollweg, K. S. & Hill, D. [Eds.]. *What is the Influence of the National Science Education Standards?: Reviewing the evidence, a workshop summary.* National Academy Press.

Wenglinsky, H. (2000). How teaching matters: Bringing the classroom back into discussions of teacher quality. (ERIC Document Reproduction Service No. ED447128).

Winking, D. (1997). Ensuring equity with alternative assessment. Retrieved May 12, 2009 from http://www.ncrel.org/sdrs/areas/issues/methods/assment/as800.html.

Zohar, A., & Dori, Y. J. (2003). Higher-order thinking skills and low-achieving students: Are they mutually exclusive? *Journal of the Learning Sciences, 12*, 145-181.