

Examining Turkish Pre-service Science Teachers' Technological Pedagogical Content Knowledge (TPACK) Based on Demographic Variables

Tezcan KARTAL¹ , Özlem AFACAN²

¹ Asst. Prof. Dr., Ahi Evran University, Kirsehir-TURKEY

² Assoc. Prof. Dr., Ahi Evran University, Kirsehir-TURKEY

Received: 14.04.2016

Revised: 15.11.2016

Accepted: 05.12.2016

The original language of article is English (v.14, n.1, March 2017, pp.1-22, doi: 10.12973/tused.10187a)

ABSTRACT

Technological pedagogical content knowledge (TPACK) can be viewed as a new concept for the educational technology world. However, there are many studies related to TPACK and TPACK cannot be considered out of context. Instead, researchers should have a deep understanding about how the results of a TPACK study may change due to its context. This study demonstrates the importance of context. The Survey of TPACK was utilized to identify 591 pre-service science teachers' (PSTs) TPACK levels and examine the validity and reliability of data obtained from 591 PSTs. Exploratory and confirmatory factor analyses were conducted for validity. After factor analyses a TPACK model with four factors (technological knowledge, content knowledge, knowledge of pedagogy, knowledge of teaching with technology) were obtained. There were, however, seven factors in the original form of the survey. This change was interpreted as the effect of the context, because the participants in the original survey and in this survey were different in terms of their teacher preparation programs and the opportunities that the teacher preparation programs provided for them. In addition, pre-service science teachers' TPACK levels were investigated on the basis of demographic variables (gender, owning computer, computer usage level and grade level). An important result obtained from the demographic variables is that pre-service science teachers' TPACK levels develop in direct proportion to their grade level. This finding supports the idea that experiences with technology and in teaching have a positive impact on TPACK.

Keywords: Pre-service Science Teachers, Technological Pedagogical Content Knowledge, TPACK Survey.

INTRODUCTION

The most important progress made in education in the last decade has been the integration of technology (Lee & Tsai, 2010). Instructional technology tools such as computers, data collecting and analyzing software, digital microscopes, hypermedia/multimedia, and interactive smart boards help students understand the nature of science and research, and obtain scientific knowledge. Using these tools in science classrooms also effectively and appropriately develops students' active participation in the process of



producing information, thinking and learning problem-solving skills (Guzey & Roehrig, 2009).

Even though opportunities to access technology are increasing, teaching practices in the classroom are not showing the expected level of improvement (Inan & Lowther, 2009; Lim & Chai, 2008). Research findings show that pre-service teachers (PSTs) and beginning teachers are using technology in an inadequate way (Tondeur, Braak, Sang, Voogt, Fisser & Ottenbreit-Leftwich, 2012). Studies related to using technology reveal that teachers have a lack of knowledge about how they can effectively integrate technology into education, and their efforts are limited in terms of content, variety and depth (Koehler, Mishra, Kereluik, Shin & Graham, 2014).

It is important to link subject-specific technology and pedagogical principles to use technology effectively in classrooms. Starting from this point of view, Mishra and Koehler (2006) extended Shulman's idea of pedagogical content knowledge (PCK) by adding technology to PCK and developed the technological pedagogical content knowledge (TPACK) framework.

The TPACK Framework

Koehler and Mishra (2005) defined technological pedagogical content knowledge as the relationships between content knowledge, technological knowledge (e.g. computer, internet, digital video) and pedagogy knowledge (practices, processes, strategies, procedures and methods of teaching and learning). Niess (2005) reported that pre-service teachers' ability to integrate technology with PCK and their views about the nature of their disciplines are important for the development of technological pedagogical content knowledge. According to Archambault and Crippen (2009), TPACK includes relationships between students, teachers, contents, and technology and its applications.

The TPACK framework consists of three main domains: knowledge of content, pedagogy and technology. Knowledge bases, represented as pedagogical content knowledge (PCK), technological content knowledge (TCK), technological pedagogical knowledge (TPK) and technological pedagogical content knowledge (TPACK), and the relationships between these knowledge bases are equally important (Koehler & Mishra, 2009).

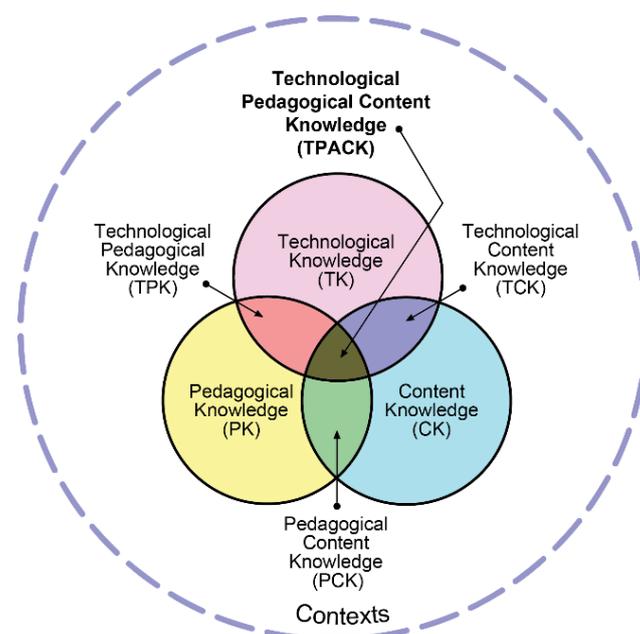


Figure 1. Components of the TPACK Framework (from tpack.org)

As seen in Figure 1, TPACK includes technology, pedagogy and content knowledge, and the intersections of these domains. The seven sub-domains of TPACK are described briefly as follows:

Technology Knowledge (TK): This includes internet, digital video, advanced technologies such as the different approaches used in providing information, as well as standard technologies like chalk, blackboard and book (Koehler, Mishra, & Yahya, 2007).

Pedagogy Knowledge (PK): Teaching and learning processes and practices including the goals of the teaching, teaching methods and strategies related to student assessment, classroom management and lesson plans are found in this sub-domain (Koehler, Mishra & Yahya, 2007; Sahin, 2011).

Content Knowledge (CK): Content knowledge is the teacher's knowledge about the subject matter that he/she teaches (Koehler & Mishra, 2009). According to Shulman (1986), content knowledge identifies knowledge and its organization in the minds of teachers.

Pedagogical Content Knowledge (PCK): PCK is the teacher's ability to transfer his/her knowledge to students in such a way as to make the content understandable to learners (Shulman, 1986; Zeidler, 2002). Shulman (1986) asserted that PCK includes teachers' understanding about what is easy or difficult in teaching specific content and that PCK distinguishes teachers from content experts.

Technological Pedagogical Knowledge (TPK): This knowledge focuses on the components and characteristics of existing technologies used in teaching and learning, and shows how particular technologies can be used in teaching (Harris & Hofer, 2011; Koehler et al., 2007).

Technological Content Knowledge (TCK): Technology and content have a deep historical relationship. Improvements in different disciplines such as medicine, history and archaeology are related to new technologies that support organizing and showing data in new and fruitful ways. For example, the viewing of X rays or the use of Carbon 14 in dating techniques is evidence of technology's effect on medicine and archaeology respectively (Koehler & Mishra, 2009). According to Harris and Hoffer (2011), TCK illuminates how to utilize technology in teaching and how to embody the basic laws of a specific content in the best way.

Technological Pedagogical Content Knowledge (TPACK): TPACK requires an understanding of how to formulate concepts and present them using technology; pedagogical techniques in teaching the subject matter while using technology in a constructive way; knowledge about what makes it difficult or easy to learn concepts and how technology engages with these issues; students' prior knowledge and epistemological/theoretical understanding; using technology to build new knowledge bases on existing ones and developing new epistemologies or strengthening the old ones (Koehler et al., 2007). Effective instruction with technology does not mean adding technology to the existing instruction and content. Instead, technology should enable new concepts to be defined and a developing sensibility of the dynamic relationships between the core elements of TPACK: Technology, pedagogy, and content (Koehler & Mishra, 2005). TPACK shows how to teach with subject-specific technologies in a way that will best support students' needs and choices, as well as embodying the topics (Harris & Hoffer, 2011).

The studies aiming to measure TPACK began with self-reported surveys about teachers' and PSTs' TPACK level (Schmidt, Baran, Thompson, Mishra, Koehler & Shin, 2009). Koehler, Shin and Mishra (2012) examined 303 publications including journal articles, and full texts of conference presentations and proceedings. They found 141 TPACK measurement tools after their revision. 31 of them were self-report measures, 20 of them were open-ended questionnaires, 31 of them were performance assessments, 30 of them were interviews and 29 of them were observations (Koehler et al., 2014).

Researchers mostly use the survey developed by Schmidt et al. (2009) (Chai, Koh, Tsai & Tan, 2011b; Koh & Chai, 2011; Koh, Chai & Tsai, 2010; Lin, Tsai, Chai & Lee, 2013). There is research that finds a factor number different from seven (Chai et al., 2011b; Koh et al., 2010), as well as research that finds a model consistent with the TPACK model that Mishra and Koehler (2006) proposed (Koh & Chai, 2011; Lin et al., 2013). Koh and Chai (2011) found seven sub-domains of TPACK with PSTs, and Lin et al. (2013) with both PSTs and in-service teachers. Contrary to these, Chai et al. (2011b) found five factors: *Technological Knowledge (TK)*, *Content Knowledge (CK)*, *Pedagogical Knowledge of Meaningful Learning (PKML)*, *Technological Pedagogical Knowledge (TPK)* and *Technological Pedagogical Content Knowledge (TPACK)*. Koh et al. (2010) studied with PSTs. They found the same number of factors as Chai et al. (2011b), but their factor constructs and names were different. They found a TPACK model with five factors: *Technological Knowledge (TK)*, *Content Knowledge (CK)*, *Knowledge of Pedagogy (KP)*, *Knowledge of Teaching with Technology (KTT)*, and *Knowledge from Critical Reflection (KCR)*. Obtaining different factors when a survey was administered to different participants can mean that factors may change due to the characteristics of participants. Mishra and Koehler (2006) reported that data from TPACK studies should not be evaluated ignoring context. An important feature of the TPACK framework is that it is not separate from outer events and effects, rather it is grounded on the context (Koehler et al., 2014). Archambault and Barnett (2010) conducted research to examine the validity of the TPACK framework. They found that it is too difficult to distinguish the sub-domains and that Technology Knowledge is the only separable knowledge base. Also, they reported that measurement of TPACK is so complex and difficult to understand because it includes inseparable components. They obtained three factors (Pedagogical Content Knowledge, Technological Curriculum-content Knowledge, and Technological knowledge) in their study.

Gaps

There are very many studies that use surveys to measure TPACK (Archambault & Crippen, 2009; Burgoyne, Graham & Sudweeks, 2010; Graham, Burgoyne, Cantrell, Smith, St Clair & Harris, 2009; Koehler & Mishra, 2005; MaKinster, Boone & Trautmann, 2010; Schmidt et al., 2009; Sahin, 2011). Most of the studies that use surveys do not refer to reliability and validity (Koehler et al., 2014). A limited number of studies have performed factor analysis for construct validity (Archambault & Barnett, 2010; Archambault & Crippen, 2009; Burgoyne, 2010; Burgoyne et al., 2010; Lee & Tsai, 2010; Lux, 2010; Schmidt et al., 2009). It is necessary to provide construct validity for the components of the model when engaging theoretical frameworks like TPACK (Archambault & Crippen, 2009). However, the relationships between demographic variables of participants and their TPACK levels have not been elaborated sufficiently (Koh et al., 2010). The most commonly used demographic variable is gender (Jang & Tsai, 2012, 2013; Koh, Chai & Tsai, 2014; Lin et al., 2013).

Examining PSTs' technology, pedagogy and content knowledge and how they link these knowledge domains is a very important research topic. More studies about the measurement of relationships between TPACK sub-domains are needed because the TPACK framework is challenging (Archambault & Barnett, 2010). In this study, PSTs' TPACK levels and their *gender*, *owning computer*, *computer usage level* and *grade level* are investigated together. It is expected that this study will give a different perspective on TPACK to the educational world in terms of different demographic variables and the model obtained from factor analysis.

Taking into account these gaps, this study aims to:

- (1) Identify PSTs' TPACK levels,
- (2) Examine the validity and reliability of the Survey of Technological Pedagogical Content Knowledge (Sahin, 2011) in a different context, and
- (3) Examine significant differences between PSTs' TPACK levels and their demographic variables.

METHODOLOGY

Cross-sectional survey research design was used in this study. This design is one of the most popular research designs used in educational studies (Christensen, Burke Johnson & Turner, 2014; Creswell, 2012). In this study, it was proposed to describe and correlate PSTs' effectiveness with regard to Technological Pedagogical Content Knowledge (TPACK) and their demographic variables. A survey (likert scale) was utilized for the purpose of collecting data. To enhance the validity of data obtained, researchers contacted PSTs one by one, used paper and pencil surveys; gave them enough time to complete the survey and helped them by explaining points with which they had difficulty (Christensen et al., 2014). All of the data collecting processes were carried out by researchers. The aim was thus to gain honest and accurate answers from the PSTs.

Participants

The participants in this study were 591 pre-service science teachers studying at a state university in a small rural city in central Anatolia. Universities in Turkey are administered by the *Council of Higher Education (CoHE)*. According to *Higher Education Statistics* there were 23 360 ($N_{\text{male}}=6\ 998$; $N_{\text{female}}=16\ 362$) pre-service science teachers in the colleges of education in 2013 (CoHE, 2013). It was assumed that participants represent PSTs across Turkey for two reasons. First, if the size of population (N) is 23 360, deviation amount (d) .05 and reliability level is .95 ($\alpha=.05$), the sample size needs to be at least 377 (Chow, Shao, & Wang, 2003; Cohen, Manion, & Morrison, 2007). Starting from this point of view, it can be said that the size of our sample ($n=591$) was enough to provide external validity (i.e. to represent the population) (Christensen et al., 2014; Levy & Lemeshow, 1999). Second, participants were determined via random sampling in which all of the individuals had the same opportunity to participate in the study and choosing one of them did not affect the others (Christensen et al., 2014). This study tried to predict the efficacies of PSTs with regard to TPACK by using multiple variables. Participants' characteristics relating to their grade and gender are given below in Table 1. About 70% of the participants were female ($N=413$). More than half of PSTs ($N=373$; 63%) owned their own computer.

Table 1. Participants' demographic variables

Department	Grade		Gender		Owning A Computer					
	N	%	N	%	Yes		No			
					N	%	N	%		
Pre-service Science Teachers	Freshman	145	25	Male	32	5	18	5	14	6
				Female	113	19	46	12	67	31
	Sophomore	141	24	Male	33	6	25	7	8	4
				Female	108	18	64	17	44	20
	Junior	173	29	Male	50	8	35	9	15	7
				Female	123	21	89	24	34	15
	Senior	132	22	Male	63	11	42	11	21	10
				Female	69	12	54	15	15	7
Total	591	100	591	100	373	100	218	100		

Data Collection Tool

The survey used for this study consisted of two sections:

The Survey of Technological Pedagogical Content Knowledge: The original form of the survey developed by Sahin (2011) consisted of 47 items and seven factors [TK (15 Items), PK (6 Items), CK (6 Items), TPK (4 Items), TCK (4 Items), PCK (7 Items) and TPACK (5 Items)]. The survey was developed through five stages: item pool, validity and reliability, discriminant validity, test-retest reliability and translation. In the beginning there were 60 items in the item pool. After views were taken from 10 experts, 47 items remained. A form including 60 items was created and experts were asked to decide on one of three options ('fully measuring', 'somewhat measuring' or 'not measuring') for each item. The survey was in the form of 5 point Likert type (1=not at all, 2=little, 3=moderate, 4=quite, and 5=complete).

The Cronbach's alpha coefficient was used for the reliability of the scale. Sahin (2011) found internal consistency scores to be .93 for TK, .90 for PK, .86 for CK, .88 for TPK, .88 for TCK, .92 for PCK and .92 for TPACK. According to the test-retest reliability scores, the reliability coefficient was found to be .80 ($p < .01$) for the TK subscale, .82 ($p < .01$) for the PK subscale, .79 ($p < .01$) for the CK subscale, .77 ($p < .01$) for the TPK subscale, .79 ($p < .01$) for the TCK subscale, .84 ($p < .01$) for the PCK subscale, and .86 ($p < .01$) for the TPACK subscale.

Personal Information Form (PIF): The PIF included information about the PSTs' gender and grade levels. In this section PSTs were also asked whether they had their own computers (Yes/No) and asked to choose one of the level options from 'beginning', 'intermediate', 'good' or 'advanced' (e.g. Muslu Kaygısız, Bağlıbel & Samancıoğlu, 2011). Gender affects the perspectives of pre-service teachers about computer usage (Teo, 2008), and the most common demographic variable examined in participants' TPACK has been gender (Archambalt & Barnett, 2010; Jang & Tsai, 2012; Koh et al., 2010; Lee & Tsai, 2010; Lin et al., 2013; Schmidt et al., 2009; Tsai, 2008). In view of the fact that participants were living in a different sociocultural structure from the participants in other studies mentioned below, it was decided that gender would be included in the Personal Information Form.

It was expected that those who had sufficient technological knowledge would be able to demonstrate an improvement in their TPACK levels (Chai, Koh & Tsai, 2010). Kartal, Kartal and Uluay (2016) found that the level of computer usage is a strong predictor of the TPACK sub-domains. For this reason, factors affecting technological knowledge such as owning computer and computer usage levels were included in the PIF (Mathews & Guarino, 2000; Robinson, 2003).

Those in the pre-service years gain experience and knowledge of their disciplines, pedagogical approaches and technological applications during their time in teacher preparation programs. Mathews and Guarino (2000) found that the teachers' year of experience had a direct effect on computer proficiency. It can be assumed that pre-service teachers' years of experience may be taken as their grade levels. So the grade level was the last demographic variable in the PIF.

Data analysis

The Statistical Package for the Social Sciences (SPSS) and the Linear Structural Relation Statistics Package (LISREL) were used in analyzing the data. All of the PSTs' answers for each item were examined and missing values were taken out of the data set before starting analysis. The reliability and validity of the survey were provided for with a step-by-step approach. Researchers decided to perform exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) with their data because they thought that the participants of the original survey and the participants of this study were different in terms of education

they had been receiving. First, EFA was performed with the SPSS. In the EFA process, the convenience of the data set for factor analysis (e.g., sample size, missing data, normality, Kaiser-Meyer-Olkin and Barlett's test of sphericity) was first examined (Field, 2009; Tabachnick & Fidell, 2013). In determining the factor numbers, principal component analysis and varimax kaiser normalization were used. The items which had eigenvalues above 1 and factor loadings equal to and above .5 were kept in the survey and items with cross loadings were omitted. Second, CFA was carried out with the help of LISREL to make sure of the survey's construct validity. CFA was seen as the natural extension of EFA (Lee, 2007). CFA aims to reveal the latent construct of the items in the survey (Brown, 2015). The findings of the survey obtained from CFA were evaluated according to different fit indexes. The evaluation began with the significance level (p) for chi-square (χ^2) and the chi-square per degree of freedom (χ^2/df). Also, the root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), root mean squares residuals (RMR), normed fit index (NFI), non-normed fit index (NNFI), comparative fit index (CFI), goodness of fit (GFI) and adjusted goodness of fit-index (AGFI) were calculated. The Item Discrimination Statistics test and Corrected Item-Total Correlation test were used for reliability. In addition, the Cronbach's alpha of each factor in the survey and correlation between factors were examined. Whether the scores of PSTs differed due to demographic variables was tested with the independent-sample t test (e.g., gender, owning computer) and One-Way Anova (e.g., computer usage level, grade level). To determine the source of significance in One-Way Anova, the Tukey test was used. In the case of a significant difference in demographic variables, the effect size was examined. Cohen's d for t test and eta-square (η^2) for One-Way Anova were calculated for the effect size of differences. Finally, multiple regression analysis was performed to examine which items predicted PSTs Knowledge of Teaching with Technology (KTT) and to what extent.

RESULTS AND DISCUSSION

Construct validity of the survey

EFA and CFA were conducted to provide the construct validity of survey. Factor analysis proposes to formulate unrelated and meaningful new factors, gathering a great number of interrelated variables (Field, 2009; Tabachnick & Fidell, 2013). The availability of the data set for factor analysis was examined in terms of (a) sample size and missing data, (b) normality, (c) linearity, (d) Kaiser-Meyer-Olkin (KMO) and Barlett's test of sphericity (BToS).

There are different views about the sufficient sample size for factor analysis. According to Comrey and Lee (1992) and Field (2009), a sample of 300 is enough to conduct factor analyses, while some researchers (Kline, 1994; Tabachnick & Fidell, 2013) find a sample of 200 enough. It is generally known that a sample size should be at least ten times the number of items and that this is important for construct validity. Taking into account all of these, we can say that the sample size was sufficient to perform factor analysis. The Kolmogorov-Smirnov (Lilliefors) test was used for normal distribution of data. According to the test results, the data had a normal distribution ($Z=.818$, $p>.05$). Linearity between variables was provided with scatter plots (Tabachnick & Fidell, 2013). KMO value was calculated as .968. For a good factor analysis, a KMO value larger than .60 is suggested (Kaiser, 1974; Sharma, 1996; Tabachnick & Fidell, 2013). The KMO value can have a value between 0 and 1, and is interpreted as 'normal' if it is between .60 and .70; as 'good' if it is between .70 and .80; as 'very good' if it is between .80 and .90; and as 'excellent' if it is equal to or larger than .90 (Field, 2009; Sharma, 1996; Tabachnick & Fidell, 2013). When taking into consideration the

results of Barlett's Test of Sphericity (Chi-square=15,619.465, df=820; $p < .001$), it was seen that data were available for factor analysis.

To determine the new factors, principal component analysis and varimax with Kaiser normalization were utilized (Field, 2009; Tabachnick & Fidell, 2013). In the EFA, 6 items (CK1, CK2, TK14, TPK1, TPK2, TCK2) with eigenvalues less than 1 (Tabachnick & Fidell, 2013), factor loadings less than .5 (e.g., Kartal et al., 2016; Koh et al., 2010; Lee & Tsai, 2010) or cross-loadings (e.g., Koh et al., 2010; Lee & Tsai, 2010; Lux, 2010) were omitted from survey. After omitting each item, factor analysis was conducted again. At the end of the EFA, the remaining 41 items formed four factors. TPK (TPK3 and TPK4), TCK (TCK2, TCK3, TCK4) and TPACK (TPACK1, TPACK2, TPACK3, TPACK4 and TPACK5) constituted the first factor; TK (except TK14) constituted the second; PK and PCK the third and CK (except CK1 and CK2) formed the fourth factor (Table 3). Koh et al. (2010) labeled the factor including TPK, TCK, and TPACK as *Knowledge of Teaching with Technology (KTT)*; the factor including PK and PCK as *Knowledge of Pedagogy (KP)*. These new factor names were used in this study (Tables 2 and 3). The final form of survey had 41 items and 4 factors (KTT, TK, KP and CK). These four factors explain 58.523% of the total variance. Eigenvalues, percentage of variance and percentage of cumulative variance values about factors are given below in Table 2.

Table 2. Total variance after rotated component matrix

Factor	Eigenvalues	Percentage of Variance (%)	Percentage of Cumulative Variance (%)
Knowledge of Teaching with Technology (KTT)	9.504	23.181	23.181
Technological Knowledge (TK)	7.619	18.583	41.764
Knowledge of Pedagogy (KP)	4.009	9.779	51.543
Content Knowledge (CK)	2.862	6.980	58.523

If the variance for each factor is less than 5%, this means that the maximum factor number is gained in explaining total variance. Henson and Robert (2006) reported that the total variance explained by factors should be at least 52%, Kline (1994) argued this value should be at least 40% and Streiner (1994) at least 50%. In similar studies (Archambault & Barnett, 2010; Kartal et al., 2016; Scherer, Wiebe, Luther & Adams, 1988; Schmidt et al., 2009), it is seen that the total variance after factor analysis is equal to and more than 50%. As a result, the total variance of this study may be seen as sufficient.

In the factor analysis, no items were omitted from the PK and PCK domains. But these domains came together and formed a new factor. This may be because PSTs could not distinguish the items. This new factor is called *Knowledge of Pedagogy* (Koh et al., 2010). Similarly, PSTs could not distinguish TPK (TPK1 and TPK2 omitted), TCK (TCK1 omitted) and TPACK items, so these three domains formed a new factor called *Knowledge of Teaching with Technology* (Koh et al., 2010). Factors and items' factor loadings are given in Table 3. When the studies that perform construct validity are reviewed, there are models with seven factors (Lin et al., 2013; Kartal, et al., 2016; Koh, Chai & Tsai, 2013; Koh et al., 2014) as well as models with different numbers of factors (Chai et al., 2011b; Chai, Koh & Tsai, 2011a; Koh et al., 2010). In the second group, the most attention-grabbing is the research of Chai et al. (2011a). Although the original TPACK framework has seven sub-domains, they find eight sub-domains in their study. They obtained these eight factors because content knowledge (CK) was divided into two parts called CKCS1 (the first teaching subject), and CKCS2 (the second teaching subject).

Table 3. Factor loading from exploratory analysis, and reliability values

	Common Factor loading	Rotated Factor loading
Factor 1 – Content Knowledge (CK, $\alpha = .817$)		
1 CK3-Following recent developments and applications in my content area	.667	.551
2 CK4-Recognizing leaders in my content area	.599	.695
3 CK5-Following up-to-date resources (ex, books, journals) in my content area	.589	.668
4 CK6-Following conferences and activities in my content area	.539	.666
Factor 2 – Technological Knowledge (TK, $\alpha = .933$)		
5 TK1-Solving a technical problem with the computer	.515	.722
6 TK2-Knowing about basic computer hardware (ex., CD-Rom, mother-board, RAM) and their functions	.546	.697
7 TK3-Knowing about basic computer software (ex., Windows, Media Player) and their functions	.563	.775
8 TK4-Following recent computer technologies	.518	.649
9 TK5-Using a word-processor program (ex., MS Word)	.575	.773
10 TK6-Using an electronic spreadsheet program (ex., MS Excel)	.529	.706
11 TK7-Communicating through Internet tools (ex., e-mail, MSN Messenger)	.541	.717
12 TK8-Using a picture editing program (ex., Paint)	.569	.738
13 TK9-Using a presentation program (ex., MS Powerpoint)	.640	.756
14 TK10-Saving data into a digital medium (ex., Flash Card, CD, DVD)	.625	.699
15 TK11-Using area-specific software	.642	.576
16 TK12-Using printer	.623	.649
17 TK13-Using projector	.600	.619
18 TK15-Using digital camera	.564	.589
Factor 3 – Knowledge of Pedagogy (KP, $\alpha = .930$)		
19 PK1-Assessing student performance	.690	.668
20 PK2-Eliminating individual differences	.675	.679
21 PK3-Using different evaluation methods and techniques	.720	.679
22 PK4-Applying different learning theories and approaches (ex, Constructivist Learning, Multiple Intelligence Theory, Project-based Teaching)	.682	.629
23 PK5-Being aware of possible student learning difficulties and misconceptions	.656	.669
24 PK6-Managing class	.610	.585
25 PCK1-Selecting appropriate and effective teaching strategies for my content area	.687	.656
26 PCK2-Developing evaluation tests and surveys in my content area	.671	.650
27 PCK3-Preparing a lesson plan including class/school-wide activities	.667	.712
28 PCK4-Meeting objectives described in my lesson plan	.691	.706
29 PCK5-Making connections among related subjects in my content area	.666	.703
30 PCK6-Making connections between my content area and other related courses	.668	.702
31 PCK7-Supporting subjects in my content area with outside (out-of-school) activities	.699	.699
Factor 4 – Knowledge of Teaching with Technology (KTT, $\alpha = .919$)		
32 TPK3-Being able to select technologies useful for my teaching career	.708	.549
33 TPK4-Evaluating appropriateness of a new technology for teaching and learning	.708	.509
34 TCK2-Using technologies helping to reach course objectives easily in my lesson plan	.738	.595
35 TCK3-Preparing a lesson plan requiring use of instructional technologies	.702	.647
36 TCK4-Developing class activities and projects involving use of instructional technologies	.702	.628
37 TPACK1-Integrating appropriate instructional methods and technologies into my content area	.720	.726
38 TPACK2-Selecting contemporary strategies and technologies helping to teach my content effective	.705	.733
39 TPACK3-Teaching successfully by combining my content, pedagogy, and technology knowledge	.690	.759
40 TPACK4-Taking a leadership role among my colleagues in the integration of content, pedagogy, and technology knowledge	.603	.666
41 TPACK5-Teaching a subject with different instructional strategies and computer applications	.680	.733

The common factor loadings of items were between .515 and .738; their rotated factor loadings were between .509 and .775. Cronbach's alpha was calculated to be .817 for Content Knowledge (CK); .933 for Technological Knowledge (TK); .930 for Knowledge of Pedagogy (KP) and .919 for Knowledge of Teaching with Technology (KTT).

The model obtained from the EFA was tested with CFA with LISREL. The sample size was sufficient for CFA (Kline, 2005; Schermelleh-Engel, Moosbrugger & Mülleer, 2003). Statistics related to the model-data fit are given in Table 4.

Table 4. Fit values and standard fit criteria

Index	Good Fit Values	Acceptable Fit Values	Survey Fit Values
χ^2	$0 \leq \chi^2 \leq 3df$	$3df < \chi^2 \leq 5df$	1,738.78
p value	$.05 \leq p \leq 1.00$	$.01 \leq p \leq .05$.048
χ^2/df	$0 \leq \chi^2/df \leq 3$	$3 < \chi^2/df \leq 5$	3.184
RMSEA	$0 \leq RMSEA \leq .05$	$.05 < RMSEA \leq .08$.038
SRMR	$0 \leq SRMR \leq .05$	$.05 < SRMR \leq .10$.044
RMR	$0 \leq RMR \leq .05$	$.05 < RMR \leq .10$.047
NFI	$.95 \leq NFI \leq 1$	$.90 < NFI < .95$.96
NNFI	$.97 \leq NNFI \leq 1$	$.95 \leq NNFI < .97$.92
CFI	$.97 \leq CFI \leq 1$	$.95 \leq CFI < .97$.94
GFI	$.95 \leq GFI \leq 1$	$.90 \leq GFI < .95$.98
AGFI	$.90 \leq AGFI \leq 1$	$.85 \leq AGFI < .90$.90

Results in Table 4 were examined in two parts. The first part used *Descriptive Measures of Overall Model Fit* and the second part used *Descriptive Measures Based on Model Comparisons*. Significant difference occurred at an acceptable level according to values ($\chi^2=1,738.78$; $p=.048$). However, it is expected that χ^2 between fit indexes will not be significant and the null hypothesis will be accepted (Schermelleh-Engel et al., 2003). Although the significance level of χ^2 was too close to the acceptable upper limit, it was seen that the p value was significant. The reason for this situation may be that χ^2 is considerably sensitive to sample size. As sample size increases, the significant level of the p value will increase. On the other hand, reducing the sample size makes χ^2 decrease and it will show a level at which model is not significant ($p>.05$). Therefore, Jöreskog and Sörbom (1993) suggested not using χ^2 alone, but instead to compare the expected value of the sample distribution (e.g., df value) with χ^2 . In this study χ^2/df was calculated as $1,738.78/546=3.184$. This value can be considered as an acceptable fit level (Kline, 2005; Schermelleh-Engel et al., 2003). Because of χ^2 value's sensibility to sample size, alternative fit indexes were examined. Good fit and acceptable fit value ranges related to these indexes are shown in Table 4 (Brown, 20015; Chermelleh-Engel et al., 2003; Jöreskog & Sörbom, 1993; Kline, 2005; Tabachnick & Fidell, 2013). Root Mean Square Error of Approximation (RMSEA) was found to be .038; Standardized Root Mean Square Residual (SRMR) to be .044 and Root Mean Squares Residual (RMR) to be .047; Normed Fit Index (NNFI) to be .92; Comparative Fit Index (CFI) to be .94; Goodness of Fit (GFI) to be .98; Adjusted Goodness of Fit-Index (AGFI) to be .90. It can be agreed that these values confirmed the model.

Reliability of the survey

To evaluate the reliability of items, item discrimination and the Corrected Item-Total Correlation test were investigated (e.g. Aydın & Kara, 2013). 27% ($n_{upper}=160$) of the participants who had the highest scores were considered as the upper group, 27% ($n_{lower}=160$) of them who had the lowest scores were considered as the lower group, and for each item the

significance of difference between the upper and lower groups was examined by t-test (Kelley, 1939). The Item Total Correlation was calculated with the Pearson Product-Moment Correlation coefficient test (Tabachnick & Fidell, 2013). According to the item analysis given in Table 5, there was no change in the number of items. When taking into consideration the items' t-test, it was clear that item correlation values were between .684 and .825 and that each item in the scale significantly discriminated the individuals belonging to the lower and upper groups ($p < .01$). Items were compatible with the whole survey and it can be assumed that they measure what they aimed to measure. The item of "Recognizing leaders in my content area" (CK, Item 2) had the highest and the item of "Using a digital camera" (TK, Item 18) had the lowest item-total correlation (Table 5).

Table 5. Means of the items, standard deviations, t-test results and item total correlations

Survey Item	Subscale	Mean	Standard Deviation	t (Up-low-27-percent group)	Corrected Item-Total Correlation
1	CK	3.079	.872	17.548 ^a	.793 ^a
2	CK	2.890	.913	16.902 ^a	.825 ^a
3	CK	2.903	.901	16.650 ^a	.822 ^a
4	CK	2.639	.879	16.264 ^a	.774 ^a
5	TK	2.543	.985	11.812 ^a	.733 ^a
6	TK	2.966	1.022	12.688 ^a	.723 ^a
7	TK	3.321	1.020	13.610 ^a	.779 ^a
8	TK	2.839	1.001	12.341 ^a	.690 ^a
9	TK	3.516	1.024	13.550 ^a	.767 ^a
10	TK	3.318	1.003	11.635 ^a	.697 ^a
11	TK	3.835	.939	13.146 ^a	.730 ^a
12	TK	3.722	.976	13.453 ^a	.751 ^a
13	TK	3.793	1.034	15.547 ^a	.794 ^a
14	TK	3.900	1.063	15.422 ^a	.754 ^a
15	TK	2.912	.991	16.097 ^a	.702 ^a
16	TK	3.225	1.156	16.438 ^a	.743 ^a
17	TK	2.939	1.158	14.787 ^a	.719 ^a
18	TK	3.477	1.122	13.620 ^a	.684 ^a
19	KP	3.138	.993	16.771 ^a	.733 ^a
20	KP	3.160	.949	16.511 ^a	.738 ^a
21	KP	3.098	.960	19.126 ^a	.770 ^a
22	KP	3.067	1.009	17.978 ^a	.745 ^a
23	KP	3.047	.921	15.769 ^a	.739 ^a
24	KP	3.287	.941	15.971 ^a	.708 ^a
25	KP	3.214	.892	16.713 ^a	.746 ^a
26	KP	3.267	.903	16.269 ^a	.735 ^a
27	KP	3.103	.962	15.842 ^a	.729 ^a
28	KP	3.148	.874	16.813 ^a	.744 ^a
29	KP	3.262	.873	17.075 ^a	.730 ^a
30	KP	3.360	.888	16.525 ^a	.732 ^a
31	KP	3.326	.926	18.358 ^a	.734 ^a
32	KTT	3.404	.883	16.539 ^a	.745 ^a
33	KTT	3.243	.864	17.150 ^a	.731 ^a
34	KTT	3.138	.915	19.804 ^a	.775 ^a
35	KTT	3.033	.978	17.545 ^a	.751 ^a
36	KTT	3.088	.938	18.301 ^a	.759 ^a
37	KTT	3.228	.924	19.106 ^a	.784 ^a
38	KTT	3.191	.929	18.149 ^a	.780 ^a
39	KTT	3.184	.959	17.894 ^a	.791 ^a
40	KTT	2.994	.975	16.208 ^a	.720 ^a
41	KTT	3.257	.919	17.332 ^a	.774 ^a

^a $p < .01$, ($N = 591$, $N_1 = N_2 = 160$)

The highest and the lowest mean scores were in Technological Knowledge (Table 5). The item for which PSTs had the highest mean is “Saving data into a digital medium (e.g., Flash Card, CD, DVD)” (TK, Item 14) ($M=3.900$; $Sd=1.603$), and the lowest was “Solving a technical problem with the computer” (TK, Item 5) ($M=2.543$; $Sd=.985$).

Correlation and regression analyses

Pearson product-moment correlation coefficient and effect size values between factors (CK, TK, KP, KTT) are given in Table 6.

Table 6. Correlations among subscales

Survey Subscales	CK	TK	KP	KTT
	(4 Items)	(14 Items)	(13 Items)	(10 Items)
		r^2	r^2	r^2
Content Knowledge (CK)	-	.469 ^a	.220	.665 ^a
Technology Knowledge (TK)		-	.538 ^a	.289
Knowledge of Pedagogical (KP)			-	.851 ^a
Knowledge of Teaching with Technological (KTT)				-
Mean	2.878	3.307	3.191	3.176
Standart Deviation	.716	.759	.685	.706

^a $p<.01$

As seen in Table 6, there was a positive correlation at a high level between KTT-CK ($r=.659$; $p\leq.001$); KTT-TK ($r=.556$; $p\leq.001$) and KTT-KP ($r=.851$; $p\leq.001$). PSTs' KTT levels had a strong correlation with CK, TK, and KP levels. Also there was a positive correlation at a high level between KP-CK ($r=.665$; $p\leq.001$); and between KP-TK ($r=.538$; $p\leq.001$). Determination values were investigated to find out the extent to which factors explain the variance of KTT and KP (Table 6). 43.4% of the total variance of KTT stemmed from CK ($r^2=.434$; large effect); 30.3% from KP ($r^2=.303$; large effect); 72.4% from KP ($r^2=.724$; large effect). However 44.2% of the total variance of PSTs' KP derived from CK ($r^2=.442$; large effect) (e.g., Cohen, 1988, 1992, 1994; Field, 2009; Muijs, 2004; Rosnow & Rosenthal, 1996). Determination values related to correlation coefficients were mainly at the 'large effect' level. Similarly to these findings, Archambault and Crippen (2009) found a low correlation between technology and content ($r=.323$; $p<.01$), and technology and pedagogy ($r=.289$, $p<.01$), but they found a high correlation between pedagogy and content ($r=.690$; $p<.01$) and they reported that the low relationships between technology and pedagogy, technology and content were to be expected. Lin et al. (2013) found high and statistically significant correlation between TPC and TCK (.83), TPK (.76), and TK (.70) among science teachers. They reported that teachers with higher self-confidence in TPACK felt the same in the factors involving technology (e.g., Technological Content Knowledge, Technological Pedagogical Knowledge, and Technological Knowledge). Also there was a comparatively weaker correlation between PCK and TPC (.30), TCK (.36), TPK (.32), and TK (.29) in science teachers.

In the study, the PSTs found themselves the most proficient in Technological Knowledge (Mean=3.307; $Sd=.759$) and the least proficient in Content Knowledge (Mean=2.878; $Sd=.716$). This correlation was especially prevalent between KTT and other factors. Chai et al. (2010) found a moderate effect between TPACK and TK, PK, and CK and determined that the PK of PSTs had an important impact on their TPACK due to the stepwise regression model on the pre-test and post-test scores. Based on this result, it can be

determined that integrating technology into teaching is related to PSTs' knowledge of pedagogy. Multiple regressions were used to determine the extent to which items of CK, TK and KP predict PSTs' KTT. The assumptions of multiple regressions (normal distribution, linearity, constant variance, absence of multiple connections between independent variables) were provided by carrying out necessary evaluations on the data set (Tabachnick & Fidell, 2013). Regression analysis results for each item that predicted PSTs' KTT levels are shown in Table 7.

Table 7. Beta and Adjusted R^2 results of regression analyses

	KTT (Beta)	Adjusted R^2	
TK	Using a presentation program (ex., MS Powerpoint)	.124 ^a	.015
	Saving data into a digital medium (ex., Flash Card, CD, DVD)	.128 ^b	.016
	Using area-specific software	.222 ^c	.049
	Using printer	.144 ^b	.021
	Using projector	.104 ^a	.011
KP	Assessing student performance	.090 ^b	.008
	Being aware of possible student learning difficulties and misconceptions	.074 ^a	.005
	Selecting appropriate and effective teaching strategies for my content area	.143 ^c	.020
	Developing evaluation tests and surveys in my content area	.142 ^c	.020
	Preparing a lesson plan including class/school-wide activities	.113 ^c	.013
	Meeting objectives described in my lesson plan	.177 ^c	.031
	Making connections among related subjects in my content area	.113 ^c	.013
	Making connections between my content area and other related courses	.075 ^a	.006
Supporting subjects in my content area with outside (out-of-school) activities	.155 ^c	.024	
CK	Recognizing leaders in my content area	.343 ^c	.118
	Recognizing leaders in my content area	.114 ^b	.013
	Following up-to-date resources (ex, books, journals) in my content area	.185 ^c	.034
	Following conferences and activities in my content area	.182 ^c	.033

^a $p < .05$, ^b $p < .01$, ^c $p < .001$

Analyses were conducted with items that formed a statistically significant difference on KTT (e.g., .05, .01 or .001 significance level). CK, KP and TK were predictors of KTT, but mainly of CK. In the CK factor the highest predictor item was "Recognizing leaders in my content area" ($\beta = .343$; $p < .001$). This item predicted 11.8% of KTT. The highest predictor in the TK factor was "Using area-specific software" ($\beta = .222$; $p < .001$) which predicted 4.9% of KTT. Finally, the highest predictor item for the KP factor was "Meeting objectives described in my lesson plan" ($\beta = .177$; $p < .001$) and this item predicted 3.1% of KTT.

Examining sub-scales according to demographic variables

Different factors formed when factor analysis were performed with developed or adopted surveys (Archambault & Barnett, 2010; Chai et al., 2010, 2011b; Koh et al., 2010; Lee & Tsai, 2010; Sahin, 2011; Schmidt et. al., 2009). The presumed reason for this situation is that the different contexts in which studies were conducted led to different factors being revealed. Context is as important as other TPACK sub-domains (TK, PK, CK, TPK, TCK, PCK, and TPACK) (Niess, 2013). Kelly (2008) indicated that the components of TPACK context were a school's philosophy and expectations; the demographic characteristics of students and teachers; teachers' knowledge, skills and disposition; cognitive, experimental, physical, psychological, social characteristics of students and teachers and the physical features of the classroom. However, a gap in the literature is the lack of analysis related to contextual components such as gender and grade level. Researchers (Kartal et al., 2016; Koh et al. 2010; Lee & Tsai, 2010) have addressed the lack of studies about relationships between

TPACK and demographic variables. With the aim of filling this gap, the PSTs' mean scores for CK, TK, KP and KTT were examined with reference to their gender, grade level, owning computer and computer usage level. The results with regard to PSTs' gender are given in Table 8.

Table 8. Results of *t*-test and Cohen's *d* according to gender variable

	Gender	N	Mean	Standard Deviation	t	p	Cohen's d
KTT	Male	178	3.280	.738	2.354	.019 ^a	.211
	Female	413	3.131	.618			
KP	Male	178	3.256	.604	1.519	.129	-
	Female	413	3.163	.716			
TK	Male	178	3.490	.806	3.882	.000 ^b	.349
	Female	413	3.229	.725			
CK	Male	178	2.931	.717	1.181	.238	-
	Female	413	2.855	.715			

^ap<.05, ^bp<.001

Even though 70% of the participants (N=413) were female, the results in TK (t=3.882; p<.05) and KTT (t=2.354; p<.001) were in favor of male participants. The effect size of this significant difference in TK (d=.349) and KTT (d=.211) was small (Cohen, 1988, 1992, 1994; Rosenthal & Rosnow, 1991). TK has the highest effect size among these two factors. Tsai (2008) suggested that males have high levels of TK because they have a more positive attitude, higher confidence and greater efficacy with regard to technology. The effect size related to the gender variable in CK and TK is small, and in KTT it is smaller. Most of the research (e.g., Archambault & Barnett, 2010; Lee & Tsai, 2010; Schmidt et al., 2009) has examined participants' gender descriptively but only a few of them examined TPACK sub-domains according to gender. Koh et al. (2010) found that males have higher scores than females in the CK, TK and KTT factors. Erdogan and Sahin (2010) reported a significant difference in TPACK, TCK, TPK, PCK and TK domains in favor of male participants.

Lin et al. (2013) compared all the sub-domains of TPACK scores with participants' (PSTs' and in-service teachers') gender. They found a statistically significant difference only for Pedagogical Knowledge and Technological Knowledge in favor of in-service male teachers. Jang and Tsai (2012) examined the effect of using interactive whiteboards (IWBs) on science and mathematics teachers' TPACK sub-domains. In their study, there was no difference in the Content Knowledge (CK), Pedagogical Content Knowledge in Context (PCKCx), Technological Knowledge (TK), and Technological Pedagogical Content knowledge in Context (TPCKCx) due to gender. Similarly, Koh and Chai (2011) did not find a significant difference in the seven sub-domains of TPACK due to gender. Jang and Tsai (2013) found a significant difference in TK and the TPACK factors they had previously obtained (Jang & Tsai, 2012) in favor of male teachers. In the study of Koh et al. (2014), related to constructivist-oriented TPACK, there were differences in the factors connected with technology such as C-TK, TCK, and C-TPACK. For these factors male teachers' awareness was higher than females' but the effect size was small.

Table 9. Results of t-test and Cohen's d according to the owning computer variable

	Owning Computer	N	Mean	Standard Deviation	t	p	Cohen's d
KTT	Yes	373	3.227	.650	2.284	.023 ^a	.195
	No	218	3.090	.788			
KP	Yes	373	3.224	.644	1.546	.123	-
	No	218	3.134	.749			
TK	Yes	373	3.484	.668	7.751	.000 ^c	.662
	No	218	3.006	.810			
CK	Yes	373	2.937	.675	2.653	.008 ^b	.227
	No	218	2.776	.772			

^ap<.05, ^bp<.01, ^cp<.001

Table 9 shows the results of the t-test and Cohen's d according to the owning computer. A majority of the participants (f=373; 63%) had their own computers. In CK (t=2.653; p<.01), TK (t=7.751; p<.001) and KTT (t=2.284; p<.05) factors, there was a significant difference in favor of having their own computer. The factor that had the highest effect size was TK (d=.662) and the effect size was at a medium level (Cohen, 1988, 1992, 1994; Rosenthal & Rosnow, 1991). Having their own computer substantially affected their levels of technological knowledge.

Table 10. One-Way Anova and Eta-Square Results related to computer usage level

	Computer Usage Level	Mean	Standard Deviation	F	p	η^2	Difference
KTT	Beginning level (N=45)(a)	2.564	.830	29.496	.000 ^e	.046	a-b
	Intermediate level (N=277)(b)	3.073	.638				a-c
	Good level (N=228)(c)	3.316	.651				a-d
	Advanced level (N=41)(d)	3.768	.652				
KP	Beginning level (N=45)(a)	2.747	.767	24.544	.000 ^e	.040	a-c
	Intermediate level (N=277)(b)	3.061	.657				b-d
	Good level (N=228)(c)	3.336	.617				
	Advanced level (N=41)(d)	3.744	.607				
TK	Beginning level (N=45)(a)	2.157	.637	216.493	.000 ^e	.137	a-b
	Intermediate level (N=277)(b)	2.993	.543				a-c
	Good level (N=228)(c)	3.708	.489				a-d
	Advanced level (N=41)(d)	4.466	.437				
CK	Beginning level (N=45)(a)	2.461	.790	18.533	.000 ^e	.035	a-c
	Intermediate level (N=277)(b)	2.760	.669				a-d
	Good level (N=228)(c)	3.012	.689				
	Advanced level (N=41)(d)	3.384	.661				

^ep<.001

Table 10 shows the relationships between PSTs' computer usage levels and the factors in the survey. As PSTs' efficacy in using computers increase, their KTT, KP, TK and CK levels also increase. The highest effect size with regards to computer usage level is in TK ($\eta^2=.137$) and it has a small size (Cohen, 1988, 1992, 1994; Rosenthal & Rosnow, 1991). The lowest effect size is in the CK factor and it has an effect lower than small. Chai et al. (2010) suggested that for the development of TPACK, PSTs should have technological knowledge at a good level. Also, Wang and Chen (2006) argued that teachers ought to have an adequate knowledge of technology to use it effectively in their classes. These two findings correspond to ours.

25% of the participants (f_M=32; f_F=113) were freshmen; 24% (f_M=33; f_F=108) were sophomores; 29% (f_M=50; f_F=123) were juniors and 22% (f_M=63; f_F=69) were seniors (Table

11). Did PSTs' grade levels lead to difference in their KTT, KP, TK and CK? The answer to this question is given with one-way Anova and eta-square (η^2) in Table 11.

Table 11. One-Way Anova and Eta-Square Results related to grade level

	Grade Level	Mean	Standard Deviation	F	p	η^2	Fark
KTT	Freshman (N=145)(a)	3.031	.878	9.004	.000 ^f	.044	a-d
	Sophomore (N=141)(b)	3.135	.76				
	Junior (N=173)(c)	3.129	.563				
	Senior (N=132)(d)	3.444	.593				
KP	Freshman (N=145)(a)	3.034	.804	9.579	.000 ^f	.047	a-c
	Sophomore (N=141)(b)	3.149	.662				
	Junior (N=173)(c)	3.157	.608				
	Senior (N=132)(d)	3.451	.594				
TK	Freshman (N=145)(a)	3.083	.876	8.985	.000 ^f	.044	a-b
	Sophomore (N=141)(b)	3.363	.695				
	Junior (N=173)(c)	3.274	.678				
	Senior (N=132)(d)	3.538	.718				
CK	Freshman (N=145)(a)	2.724	.839	5.487	.001 ^e	.027	a-b
	Sophomore (N=141)(b)	2.822	.676				
	Junior (N=173)(c)	2.917	.623				
	Senior (N=132)(d)	3.054	.690				

^ep<.01, ^fp<.001

As PSTs' grade levels increased, their mean scores for KTT (F=9.004; p<.001), KP (F=9.579; p<.001), TK (F=8.985; p<.001), and CK (F=5.487; p<.001) also increased (Table 11). An updated science teacher education program has been carried out in Turkey since 2006. This program follows a 4-year period (fall and spring terms in each year). In the first two years content courses (physics, chemistry, biology) are dominant, as well as a small number of pedagogy courses. Method courses related to technology use in education start in the third year. In the last year PSTs go into their practicum and teach with their practice teachers. Within this context, the source of the difference in PSTs' TK according to their year of study may be courses related to technology and technology-pedagogy in the second and third years of the program. The difference between KTT and KP is mainly among freshmen and seniors. KP and KTT levels in particular go up in the final years, because in the last years of their program PSTs are provided with the professional experience in which to integrate or transfer pedagogy, technology and content knowledge. Niess, Suharwoto, Lee and Sadri (2006) found that naïve teachers with low pedagogic efficacy are less capable of linking technology, pedagogy and content knowledge; Pierson (2001) obtained the result that even if teachers with a low pedagogical knowledge understand technology well, they have difficulty in connecting pedagogy and technology. It is expected that these results will clarify why our results are in favor of seniors.

CONCLUSION

The original form of the survey consisted of 47 items and 7 factors (TK, PK, CK, TPK, TCK, PCK, and TPACK). In the EFA process 6 items were omitted and the survey included 4 factors. With CFA, the aim was to test the model obtained from EFA. After EFA, PSTs' TK and CK each constituted a distinct factor, but other domains merged into the KP (Knowledge of Pedagogy) and KTT (Knowledge of Teaching with Technology) factors. Although TK, CK, PK, TCK, TPK, PCK and TPACK domains are distinct in the original form, in this study these domains could not be separated. PK and PCK domains came together in the KP factor;

TPK, TCK and TPACK in the KTT factor (Koh et al., 2010). Similarly, TPACK surveys reveal different models with different factors when performed with different participants. Therefore, more studies about PSTs' TPACK should be performed; if different factors emerge these differences should be investigated, so the epistemological knowledge of PSTs can be approached from different perspectives. This study shows that teacher preparation programs provide limited experiences about teaching and learning with technology to pre-service teachers. PSTs failed to distinguish the subdomains. Only TK and CK emerged as separate factors. The others fell into two groups: Teaching and learning processes without technology formed KP; teaching and learning processes with technology formed KTT. PSTs seemed to distinguish only between the inclusion or exclusion of technology.

There was positive correlation between each factor (CK, TK, KP and KTT). The strongest correlation was between KP and KTT. It is expected that when the levels of PSTs' KP are high their KTT levels will also be high. The KTT levels of PSTs are predicted by the items in the CK, TK and KP. The highest predictor item is in the CK factor. With regard to gender, TK and KTT scores are in the favor of males. PSTs who owned a personal computer attained higher scores in TK, CK, and KTT. The computer usage levels showed significant differences in all of the factors. This difference was in favor of PSTs who could use a computer at an advanced level. In addition, as the participants' grade levels increased, their scores in all factors differed significantly and developed.

LIMITATIONS AND IMPLICATIONS

We wish the following implications of this study to be used as a guide for future research: Firstly, when a TPACK survey was administered to a different study group from the original group with which the survey was developed, structural differences were seen in the survey. The reason for these differences is presumed to be that PSTs' TPACK differs according to the context in which the study is carried out. It can be said that PSTs' schools, the regions they live in, the opportunities that their teacher preparation program provides for them (e.g., technological software and hardware, smart board, class size, equality of opportunity) and PSTs' individual differences (e.g., psychological, social, motivational and in terms of teaching experience) all have impacts on their TPACK. This study is different from others in many aspects. We investigated to what extent CK, TK, and KP factors predicted PSTs' KTT levels, and the differences in PSTs' CK, TK, KP and KTT levels due to demographic variables.

This study was limited to the survey as a data-collecting tool. The findings were obtained within the frame of PSTs' answers related to their personal views and the characteristics of the population. By using different methodologies such as individual or focus group discussion, examining lesson plans and observing behaviors in the class, more detailed results could be derived. For example, PSTs' lesson plans could be incorporated in the data-collecting process to describe their TPACK level (Lyublinskaya & Tournaki, 2014). Interviews of participants enrich the data collecting process. Also, longitudinal studies in which different research designs are used together could be conducted to determine whether these measurements will predict to what extent PSTs will be successful in integrating technology when they start teaching in real classrooms (Abbitt, 2011).

This study was carried out with pre-service science teachers alone and only the CK, TK, KP, KTT factors were brought to light. Conducting this survey in different disciplines might lead to different factors. These different factors and results could provide the opportunity to better understand the complex relationships between technology, pedagogy and content knowledge. Also, there was no data about PSTs' prior knowledge and experiences of teaching with technology. In future research, after the workshops in which pre-service teachers are

introduced to the technological software and hardware they can use in their teaching, their TPACK levels could be examined.

REFERENCES

- Abbitt, J. T. (2011). Measuring technological pedagogical content knowledge in preservice teacher education: A review of current methods and instruments. *Journal of Research on Technology in Education*, 43(4), 281-300.
- Archambault, L. M., & Barnett, J. H. (2010). Revisiting technological pedagogical content knowledge: Exploring the TPACK framework. *Computers & Education*, 55(4), 1656-1662.
- Archambault, L. M., & Crippen, K. (2009). Examining TPACK among K-12 online distance educators in the United States. *Contemporary Issues in Technology and Teacher Education*, 9(1), 71-88.
- Aydın, F., & Kara, F. N. (2013). Öğretmen Adaylarının Teknolojiye Yönelik Tutumları: Ölçek Geliştirme Çalışması. *Türk Fen Eğitimi Dergisi (TUFED)*, 10(4), 103-118.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York: Guilford Publications.
- Burgoyne, N. (2010). *Investigating the reliability and construct validity of a measure of preservice teachers' self-efficacy for TPACK*. Unpublished master's thesis. Provo, UT: Brigham Young University.
- Burgoyne, N., Graham, C. R., & Sudweeks, R. (2010, March). The validation of an instrument measuring TPACK. In *Society for Information Technology & Teacher Education International Conference* (Vol. 2010, No. 1, pp. 3787-3794).
- Chai, C. S., Koh, J. H. L., & Tsai, C. C. (2010). Facilitating Preservice Teachers' Development of Technological, Pedagogical, and Content Knowledge (TPACK). *Educational Technology & Society*, 13(4), 63-73.
- Chai, C. S., Koh, J. H. L., & Tsai, C. C. (2011a). Exploring the factor structure of the constructs of technological, pedagogical, content knowledge (TPACK). *The Asia-Pacific Education Researcher*, 20(3), 595-603.
- Chai, C. S., Koh, J. H. L., Tsai, C. C., & Tan, L. L. W. (2011b). Modeling primary school pre-service teachers' Technological Pedagogical Content Knowledge (TPACK) for meaningful learning with information and communication technology (ICT). *Computers & Education*, 57(1), 1184-1193.
- Chow, S. C., Shao, J., & Wang, H. (2003). *Sample Size Calculations In Clinical*. New York: Marcel Dekker.
- Christensen, L. B., Burke Johnson, R., & Turner, L. A. (2014). *Research methods, design, and analysis* (12th Ed.). Boston: Pearson education, Inc.
- Cohen, J. C. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cohen, J. C. (1992). Statistical power analysis. *Current Directions in Psychological Science*, 1(3), 98-101.
- Cohen, J. C. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (6th Ed.). London and New York: Taylor & Francis Group.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.
- Council of Higher Education (CoHE). (2013). *Instructional Year Higher Education Statistics* <http://www.osym.gov.tr/dosya/1-69402/h/13ogretimalanlisansogrencisay.pdf> Accessed: 05 August 2014.

- Creswell, J. W. (2012). *Educational research: planning, conducting, and evaluating quantitative and qualitative research* (4th Ed.). Boston: Pearson Education, Inc.
- Erdogan, A., & Sahin, I. (2010). Relationship between math teacher candidates' Technological Pedagogical And Content Knowledge (TPACK) and achievement levels. *Procedia-Social and Behavioral Sciences*, 2(2), 2707-2711.
- Field, A. (2009). *Discovering statistics using SPSS (and sex and drugs and rock 'n' roll)* (3rd Ed.). Los Angeles, CA: SAGE Publication.
- Graham, R. C., Burgoyne, N., Cantrell, P., Smith, L., St Clair, L., & Harris, R. (2009). Measuring the TPACK confidence of inservice science teachers. *TechTrends*, 53(5), 70-79.
- Guzey, S. S., & Roehrig, G. H. (2009). Teaching Science with Technology: Case Studies of Science Teachers' Development of Technological Pedagogical Content Knowledge (TPCK). *Contemporary Issues in Technology and Teacher Education*, 9(1), 25-45.
- Harris, J. B., & Hofer, M. J. (2011). Technological pedagogical content knowledge in action: A descriptive study of secondary teachers' curriculum-based, technology-related instructional planning. *Journal of Research on Technology in Education*, 43(3), 211-229.
- Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research. *Educational and Psychological Measurement*, 66(3), 393-416.
- Inan, F. A., & Lowther, D. L. (2010). Factors affecting technology integration in K-12 classrooms: A path model. *Educational Technology Research and Development*, 58(2), 137-154.
- Jang, S. J., & Tsai, M. F. (2012). Exploring the TPACK of Taiwanese elementary mathematics and science teachers with respect to use of interactive whiteboards. *Computers & Education*, 59(2), 327-338.
- Jang, S. J., & Tsai, M. F. (2013). Exploring the TPACK of Taiwanese secondary school science teachers using a new contextualized TPACK model. *Australasian Journal of Educational Technology*, 29(4), 566-580.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software International.
- Kaiser, H.F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- Kartal, T., Kartal, B., & Uluay, G. (2016). Technological pedagogical content knowledge self-assessment scale (TPACK-SAS) for pre-service teachers: Development, validity and reliability. *International Journal of Eurasia Social Sciences*, 7(23), 1-36.
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30(1), 17.
- Kelly, M. A. (2008). Bridging digital and cultural divides: TPCK for equity of access to technology. *The handbook of technological pedagogical content knowledge (TPCK) for educators*, 31-58.
- Kline, P. (1994). *An easy guide to factor analysis*. New York, NY: Routledge.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd Ed.). New York: A Division of Guilford Publications, Inc.
- Koehler, M. J., & Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of educational computing research*, 32(2), 131-152.
- Koehler, M. J., & Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research*, 32(2), 131-152.
- Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)?. *Contemporary issues in technology and teacher education*, 9(1), 60-70.

- Koehler, M. J., Mishra, P., & Yahya, K. (2007). Tracing the development of teacher knowledge in a design seminar: Integrating content, pedagogy and technology. *Computers & Education*, 49(3), 740-762.
- Koehler, M. J., Mishra, P., Kereluik, K., Shin, T. S., & Graham, C. R. (2014). The technological pedagogical content knowledge framework. In *Handbook of research on educational communications and technology* (pp. 101-111). New York: Springer.
- Koehler, M. J., Shin, T. S., & Mishra, P. (2012). How do we measure TPACK? Let me count the ways. *Educational technology, teacher knowledge, and classroom impact: A research handbook on frameworks and approaches*, 16-31.
- Koh, J. H. L., & Chai, C. S. (2011). Modeling pre-service teachers' technological pedagogical content knowledge (TPACK) perceptions: The influence of demographic factors and TPACK constructs. In G. Williams, P. Statham, N. Brown & B. Cleland (Eds.), *Changing Demands, Changing Directions. Proceedings ascilite Hobart 2011* (pp. 735-746). <http://www.ascilite.org.au/conferences/hobart11/procs/Koh-full.pdf>
- Koh, J. H. L., Chai, C. S., & Tsai, C. C. (2010). Examining the technological pedagogical content knowledge of Singapore pre-service teachers with a large-scale survey. *Journal of Computer Assisted Learning*, 26(6), 563-573.
- Koh, J. H. L., Chai, C. S., & Tsai, C. C. (2013). Examining practicing teachers' perceptions of technological pedagogical content knowledge (TPACK) pathways: A structural equation modeling approach. *Instructional Science*, 41(4), 793-809.
- Koh, J. H. L., Chai, C. S., & Tsai, C. C. (2014). Demographic Factors, TPACK Constructs, and Teachers' Perceptions of Constructivist-Oriented TPACK. *Educational Technology & Society*, 17(1), 185-196.
- Lee, M. H., & Tsai, C. C. (2010). Exploring teachers' perceived self-efficacy and technological pedagogical content knowledge with respect to educational use of the World Wide Web. *Instructional Science*, 38(1), 1-21.
- Lee, S. Y. (2007). *Structural equation modeling: A bayesian approach* (Vol. 711). John Wiley & Sons.
- Levy, P. S., & Lemeshow, S. (1999). *Sampling of Populations: Methods and Applications* (3rd Ed.). Wiley Series in Survey Methodology, New York: A Wiley-Interscience Publication.
- Lim, C. P., & Chai, C. S. (2008). Teachers' pedagogical beliefs and their planning and conduct of computer-mediated classroom lessons. *British Journal of Educational Technology*, 39(5), 807-828.
- Lin, T. C., Tsai, C. C., Chai, C. S., & Lee, M. H. (2013). Identifying science teachers' perceptions of technological pedagogical and content knowledge (TPACK). *Journal of Science Education and Technology*, 22(3), 325-336.
- Lux, N. J. (2010). *Assessing technological pedagogical content knowledge*. Unpublished doctoral dissertation. Boston, MA: Boston University School of Education.
- Lyublinskaya, I., & Tournaki, N. (2014). Preparing Special Education Teachers for Teaching Mathematics and Science with Technology by Integrating the TPACK Framework into the Curriculum: A Study of Teachers' TPACK Development through Assessment of Lesson Plans. *Journal of Technology and Teacher Education*, 22(4), 449-470.
- MaKinster, J., Boone, W., & Trautmann, N. (2010). Development of an instrument to assess science teachers' Perceived Technological Pedagogical Content Knowledge. In *Poster presented at the 2010 Annual Conference of the National Association for Research in Science Teaching, Philadelphia, PA*.
- Mathews, J. G., & Guarino, A. J. (2000). Predicting teacher computer use: A path analysis. *International Journal of Instructional Media*, 27(4), 385-392.

- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *The Teachers College Record*, 108(6), 1017-1054.
- Muijs, D. (2004). *Doing quantitative research in education with SPSS*. California: Sage Publications.
- Muslu Kaygısız, G., Bağlıbel, M., & Samancıoğlu, M. (2011). Computer Use of Science Teachers and Their Attitudes toward Computer Supported Teaching: A Sample from Turkey. *Turkish Science Education (TUSED)*, 8(2), 78-90.
- Niess M. L., Suharwoto G., Lee K., & Sadri P. (2006) *Guiding inservice mathematics teachers in developing Technology Pedagogical Content Knowledge (TPCK)*. Paper presented at the Society for Information Technology and Teacher Education International Conference, Chesapeake, VA.
- Niess, M. L. (2005). Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. *Teaching and teacher education*, 21(5), 509-523.
- Niess, M. L. (2013). Central component descriptors for levels of technological pedagogical content knowledge. *Journal of Educational Computing Research*, 48(2), 173-198.
- Pierson, M. E. (2001). Technology integration practice as a function of pedagogical expertise. *Journal of Research on Computing in Education*, 33(4), 413-430.
- Robinson, W. I. (2003). *External, and internal factors which predict teachers' computer usage in K-12 classrooms*. Detroit, MI: Wayne State University.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd Ed.). New York: McGraw Hill.
- Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods*, 1, 331-340.
- Sahin, İ. (2011). Development of survey of technological pedagogical and content knowledge (TPACK). *TOJET: The Turkish Online Journal of Educational Technology*, 10(1), 97-105.
- Scherer, R. F., Wiebe, F. A., Luther, D. C., & Adams, J. S. (1988). Dimensionality of coping: Factor stability using the ways of coping questionnaire. *Psychological Reports*, 62(3), 763-770.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74.
- Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK) the development and validation of an assessment instrument for preservice teachers. *Journal of Research on Technology in Education*, 42(2), 123-149.
- Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Wiley & Sons, Inc.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Streiner, D. L. (1994). Figuring out factors: The use and misuse of factor analysis. *Canadian Journal of Psychiatry*. 39(3), 135-140.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th Ed.). New York: Pearson Education.
- Teo T. (2008) Pre-service teachers' attitudes towards computer use: a Singapore survey. *Australasian Journal of Educational Technology*, 24, 413-424.
- Tondeur, J., Van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education*, 59(1), 134-144.

- Tsai, C. C. (2008). The preferences toward constructivist Internet-based learning environments among university students in Taiwan. *Computers in Human Behavior, 24*(1), 16-31.
- Wang, Y. M., & Chen, V. D. T. (2006). Untangling the confounding perceptions regarding the stand alone IT course. *Journal of Educational Technology Systems, 35*(2), 133-150.
- Zeidler, D. L. (2002). Dancing with maggots and Saints: Visions for Subject Matter Knowledge, Pedagogical Knowledge, and Pedagogical Content Knowledge in Science Teacher Education Reform. *Journal of Science Teacher Education, 13*(1), 27-42.

Examine middle school students' constructivist environment perceptions in Turkey: School location and class size

Nevzat YİĞİT¹, Muhammet Mustafa ALPASLAN² , Yasin CINEMRE³, Bilal BALCIN³

¹ Associate Prof. Dr., KTÜ Fatih Education Faculty, Department of Primary Education, Trabzon-TURKEY

² Assistant Prof. Dr., Muğla Sıtkı Koçman University, Mugla-TURKEY

³ Teacher, Ministry of National Education, Trabzon-TURKEY

Received: 14.06.2016

Revised: 22.10.2016

Accepted: 20.11.2016

The original language of article is English (v.14, n.1, March 2017, pp.23-34, doi: 10.12973/tused.10188a)

ABSTRACT

This study aims to examine the middle school students' perceptions of the classroom learning environment in the science course in Turkey in terms of school location and class size. In the study the Assessing of Constructivist Learning Environment (ACLE) questionnaire was utilized to map students' perceptions of the classroom learning environment. The sample included 1882 students from Grades 6-8 randomly selected twenty schools in a northern province in Turkey. Data analysis revealed that there were significant differences between the perceptions of students from rural/urban schools and small/large classes. More specifically, students in small classes and rural schools reported their classroom learning environments more positively than those who were in large classes and urban schools in dimensions of thought provoking, collaboration, life relevance, concurrent learning and assessment, and bringing different viewpoints. Implications and future directions were discussed.

Keywords: Class size, constructivism, classroom learning environment, school location.

INTRODUCTION

The classroom learning environment is defined as a social atmosphere in which learning occurs (Fraser, 2007). Students' achievements and behaviors are influenced by factors in the classroom (Khalil & Saar, 2009). In such a learning environment that every psychological and sociological learning tool is important in learning process, it is undeniable that learners' reflection to their teaching-learning environment can be an indicator of student outcomes at schools. Moreover, Fraser (2014) underscores because students spend nearly 20,000 hours in classrooms until graduating from colleges, students' feelings on classroom learning environments influence educational desired outcomes and thus, the research addressing learning environments should be taken as considerable of importance. Therefore, in last decades classroom learning environment research on teaching-learning has drawn attentions of many researchers in science education (Fraser, 2012; Khalil & Saar, 2009; Liu, 2010).

Along with recent changes on our understanding of how learning occurs, educational reforms in science curriculum, which are dominated by constructivist paradigm, have been



recently launched in many countries including Turkey (Cil, 2015; Taber, 2008). Apart from the behaviourism-, learning occurs by solely transmitting knowledge from one to another one- the constructivist paradigm pays attentions into the connections of new knowledge with learners' experience in the learning process. Additionally, the constructivist paradigm highlights the importance of cognitive and environmental factors in the learning process. Because the learning occurs in classrooms, the classroom learning environments should be designed accordingly. Thus, the particular focus of this study was given on examining students' perceptions of classroom learning environment in terms of the constructivist paradigm. Bektas and Taber (2009) state that the constructivist classrooms are where the learners work together, present their ideas, take active roles and are assessed in the learning process. Addition to these, rather than transmitting new knowledge, constructivist classrooms should promote students to evaluate their existing knowledge with new one (Evrekli, Inel, Balim, & Kesercioglu, 2009). Therefore, in order to cultivate the meaningful learning in science classrooms, it is important to investigate to what extent students evaluate their learning environment as constructivist and factors that might influence on their perceptions of the classroom learning environment.

Research has associated students' perceptions of the classroom learning environment with a broad range of variables in educational research including students' motivation, attitudes and achievement, teachers' job satisfaction, and so on. The classroom learning environment should motivate students towards learning science, and be designed to help them pay more attention scientific phenomena. Research has demonstrated that students' perceptions of classroom learning environments are positively related to attitudes towards science (den Brok, Telli, Cakiroglu, Taconis, & Tekkaya, 2010), motivation (Könings, Brand-Gruwel, & van Merriënboer, 2011), and achievement goal orientation (Ghanizadeh, & Jahedizadeh, 2015). Furthermore, it positively influences teachers' job satisfaction with their profession (Tillman & Tillman, 2008). For example, Ozkal, Tekkaya, Cakiroglu, and Sungur (2009) studied the influence of students' perceptions of the classroom learning environment (personal relevance, uncertainty, critical voice, shared control, and student negotiation) on their epistemological beliefs (tentative and fixed beliefs) and learning strategies in science. A total of 1152 eight grade students participated in their study. Ozkal et al found that all dimensions of students' perceptions of the classroom learning environment directly influenced students' tentative and fixed beliefs in science (e.g., $\beta = 0.12$, $p < 0.01$ for critical voice and tentative beliefs). Furthermore, students' perceptions of the classroom learning environments perceptions directly and indirectly contributed to their learning approach (e.g., $\beta = 0.12$, $p < 0.01$ for both student negotiation and learning strategies). Their study demonstrated that how students felt about their classroom learning environments led them to construct their ideas about the nature of knowledge and knowing and choose strategies for learning. This highlights the importance of fostering students' perceptions of learning environment to develop the other desired outcomes for educational purposes.

According to Fraser (2007), the classroom learning environment research was originated from Walberg's and Moss' work during 1970s. In the theory of Moss (1979) about the classroom learning environment, the classroom learning environment consisted of the environmental and the personal systems that interaction between these systems influences each other. The environmental system includes physical setting, organizational factors, and humane aggregate and social climate (e.g. school location; Liu, 2010). The personal system includes sociodemographic variables, expectations, personality factors, and coping skills (e.g. students' gender; Liu, 2010). In the present study, we focus on two environmental system variables including class size and school location, in which may potentially affect students' perceptions of learning environment.

Class size

There is a huge discussion on whether class size should be small or large for economic and educational purposes (Harfitt, 2015; Graue, Hatch, Rao, & Oen, 2007). According to Fabunmi, Brai-Abu, and Adeniji (2007), class size was a crucial determinant on student's academic performance in Nigeria, and class congestion had negative impact on both teacher productivity and student learning input. However, Slavin (1990) reviewed research on the class size and argued that class size had positive effects on student outcomes but this effect on the achievement was extremely small. Addition to this, Glass and his colleagues (1982) published a meta-analysis combining the results of 77 empirical studies pertaining to the relationship between class size and achievement, and found that small classes were associated with higher achievement at all grade levels. It was found that the major benefits of reducing class size occurred where the number of students in the class was fewer than 20 (Glass et al., 1982; U. S. Department of Education, 1999).

In the literature few studies have focused on the influence of class size on students' perceptions of the classroom learning environment. A study by den Brok and his colleagues (2006) investigated the association between learning environment and the constructivist perceptions of a sample of 665 middle school students in California, and revealed that students in the larger class perceived constructivist environment as more positively than did in the smaller class although no significant difference was found between students in large and small classrooms. In another study, Levy, den Brok, Wubbels and Brekelmans (2003) examined the effects of school-, class- and teacher-level variables on the students' perceptions of the classroom learning environment. Levy et al. reported that class size was negatively associated with students' perceptions of the classroom learning environment. More specifically, the smaller the class size is, the more students perceived their teachers' behavior positively including satisfaction, friendly, and leadership. We have not located any study determining the influence of the class size on students' perceptions of the classroom learning environment amongst middle school students in Turkey. Furthermore, these two studies above demonstrate the direction (negative or positive) of the influence of class size on the classroom learning environment may vary one school from another. Thus, there is a need for further studies into the influence of class size on the classroom learning environment in Turkey.

School location

Differences between urban/rural area students' perceptions of the classroom learning environment were associated with broad variables by educational researchers. Research in urban/rural differences, in a broad sense, has focused on the differences in achievement, appropriateness of urban/rural achievement measures, and accessibility to resources (Young, 1998; Barton, 2007). However, according to Barton (2007), it is complex to compare student achievement relative to location as a careful consideration must be given to how individual factors interrelate with other factors, (e.g. classroom, school) to compound larger effect size (Norman, 2001; Barton, 2007). A study by Young (1998) revealed that the location of the school had a significant effect on student achievement, that students who attended in urban schools performed better than those in rural schools. Another study (Ballou & Podgursky, 1995) examining the difference between teachers' perception rural and urban schools reported that teachers in rural schools perceived the school environment were more positively than did in urban schools. In the literature, studies on students' constructivist perceptions in rural/urban locations have mostly focused only urban or rural school students. We have not found any comparison study dealing the constructivist classroom learning environment in urban and rural schools. Therefore, this study will address this gap and help future researchers have a closer look at the complexity mentioned by Norman et al. (2001) in the learning process and differences between rural/urban area's student perceptions.

The current study

Ministry of National Education of Turkey (MONE) introduced a new science education curriculum so as to improve science education from Grade 4 through Grade 8 in 2004 and implemented it in 2005 (MONE, 2005). With this reform movement, it was purposed to change science teaching from teacher-oriented to student-oriented. Several reasons can be account for this reform. One was to aim at catching the international level of constructivist oriented science teaching by giving students active roles in learning, rather than giving them passive roles in learning science, and meeting the requirement of European Union to enter in it (MONE, 2005). The other was that it was thought that this reform would be cure for unsatisfactory score in the international project and increase students' attitudes towards science. In the Third International Mathematics and Science Study (TIMSS) in 1999, for example, Turkey ranked 33rd out of 38 countries in Grade 8 level test showed that Turkish students' attitudes towards science were unsatisfactory (Den Brok et al., 2010). However, in TIMSS 2007 at Grade 8 level Turkey increased 21 points the average of students and ranked 31st out of 59 countries (Bayraktar, 2010).

Although it seems that reform movement gave positive results in students' achievement in international science tests, it is still unsatisfactory. One reason can be that the reform movement directly addressed the curriculum, rather than being purposed teachers' behavior and reflection on their teaching (Den Brok et al., 2010; Ozata-Yucel & Ozkan, 2014). Considering the requirement of the constructivist learning environment, teachers' reflecting on their teaching might be improved in the direction of less teacher-centered and more practice-oriented lessons that stimulate students' attitudes towards science (Den Brok et al., 2010; Yasar & Sozbilir, 2013). Another reason can be students' medium attitudes towards science. Cokadar and Kulce (2008), for instance, found a sample of 503 middle school students had medium attitudes towards science with 2.5% of the students showing (strongly) negative attitudes towards science. Therefore, there is still a need for deeper insight into the factors that might lead students to have negative attitudes towards science in Turkey.

Students' perceptions of the classroom learning environment are the strongest predictor of student attitudes toward to learning science (Cannon, 1997). A classroom learning environment should help teachers teach better, and consequently give students opportunities to learn better (Hassen & Childs, 1998; Liu, 2010). In the light of the literature, the purpose of this study is twofold: a) to determine differences between in students' perceptions of the classroom learning environment in large and small classrooms, and b) to examine any differences between students' perception of their classroom learning environment in rural and urban areas in terms of the constructivist paradigm. With this purpose, the following research questions were sought to address:

- Does school location (urban and rural) influence students' perceptions of the constructivist classroom learning environments?
- Does class size (small and large) influence students' perceptions of the constructivist classroom learning environments?

METHODS

A self-reported survey design was utilized to address the research questions. The Assessing of Constructivist Learning Environment (ACLE), as a self-report questionnaire, was used to collect data. Multivariate analysis of variance (MANOVA) was utilized to identify school location and class size differences in students' perceptions of the constructivist classroom learning environments.

a) Participants

The research sample consisted of 1882 Grade 6, Grade 7 and Grade 8 students from twenty middle schools in a northern city in Turkey. Table 1 represents the descriptive characteristic for the research sample. Glass and his colleagues (1982) in their meta-analysis study reported the ideal class size as 20 students. Therefore, in this study we identified a class as large if the number of students in that class was more than 20 students and as small if the number of students in that class was equal or less than 20 students and small. As for school location (urban or rural), we used the predetermined classification made by the Ministry of Turkish Interior, and randomly selected ten schools from each area.

Data were collected in May 2011. All instruments were administrated to the students in regular class hour under the supervision of their science teacher. Students were given 30 minutes to complete the instruments.

Table 1. *Distribution of participants in the study*

	School location		Class size		Total
	Rural	Urban	Small (n≤20)	Large (20<n)	
Gender					
Girls	266	649	255	660	915
Boys	279	688	272	695	967
Grade level					
Grade 6	173	474	185	462	647
Grade 7	178	443	157	464	621
Grade 8	194	420	185	429	614
Total	545	1337	527	1355	1882

n: the number of students in a classroom

b) Instrumentation

In this study the ACLE, developed by Arkun and Askar (2010), was used to examine students' perceptions of the constructivist classroom learning environment. Although in the classroom learning environment literature a various self-measurement instruments are available (e.g., Constructivist Learning Environment Survey (CLES) by Taylor and his colleagues, 1997), there are several reasons for selecting this particular instrument. First, as a self-report questionnaire, the ACLE was developed by native Turkish speaker researchers. This enabled the developers to consider the cultural dynamics of potential participants in process of developing items in the ACLE. Instruments developed and validated in other cultures, and then translated into Turkish may not be similar with Turkish cultural context (Hofer, 2008). Lastly, the ACLE nicely captures the aspects of constructivist classrooms including concurrent assessment in learning process highlighted by researchers (e.g., Bektas & Taber, 2009), and by the Ministry of National Education (MoNE, 2005).

The ACLE was based on the constructivist paradigm that supports recent research in science education. It includes relevant dimensions from past classroom learning environment questionnaires and combines these with dimensions that measure particular aspects of constructivism (Arkun & Askar, 2010). The ACLE is intended to map students' constructivist perceptions in six dimensions; *student-centered*, *thought provoking*, *collaborative*, *life relevant*, *concurrent learning and assessing*, and *different view-points*. Student-centered scale captures whether students play an active role in learning or not. Thought provoking scale measures to what degree learning environment prepossess learners. Collaborative scale captures to what degree learning environment allows for social interaction that is an important element of constructing knowledge. Life relevance scale is about knowledge is constructed by experience. Concurrent learning and assessment measures to what degree environment enable

learners to monitor their learning. Bringing different viewpoints captures to what degree learning environment helps students develop their view points (Arkun and Askar, 2010).

The ACLE consists of a set of 28 items, which is on a 7-point Likert scale the degree to which they rate their constructivist perceptions (1 low to 7 high). Because a 7-point scale may be confusing for middle school aged students, we decided to convert the ACLE into a 5-point Likert (1= strongly disagree... 5= strongly agree) scale. To assess the ACLE's validity and dimensionality, Confirmatory Factor Analysis (CFA) was conducted with AMOS 18. Hu and Bentler (1999) recommended $CFI > .95$ or $RMSEA < .06$ for good fit and $CFI > .90$ and $RMSEA < .08$ for moderate fit. The initial CFA results were not in a satisfactory level ($\chi^2 (335, N=1882) = 1186.13, p < .001, SRMR = .062, RMSEA = .065, CFI = .89$). It was suggested to use cut-off values of .30 for factor loading (Cabrera-Nguyen, 2010). Thus, Item 7 and Item 21 were removed because of low factor loading (.17 and .28, respectively). The CFA was re-run and those results indicated a good model fit ($\chi^2 (284, N = 1882) = 960.51, p < .001, SRMR = .03, RMSEA = .036, CFI = .95$). Arkun and Askar (2010) reported ACLE's reliability with 247 college students as .76 for student-centered, .88 for thought provoking, .75 for collaborative, .89 for life relevance, .81 for concurrent learning and assessment, and .83 for bringing different viewpoints. In this study, we found ACLE's reliabilities as .72 for student-centered, .78 for thought provoking, .70 for collaborative, .69 for life relevance, .84 for concurrent learning and assessment, and .83 for bringing different viewpoints. ACLE's reliability values were in acceptable range, which were around .70 (George & Mallery, 2003). An example item from each scale in ACLE was presented in Table 2.

Table 2. Typical item and reliability for each ACLE scale in the present study

Scale	Typical item	Reliability	# of items
Student-centered	I have opportunity to decide about what subject I will work on	.72	4
Thought- Provoking	I am promoted to think in lesson	.78	7
Collaborative	I don't hesitate share my idea with others	.70	3
Life Relevance	I believe what I learn in lesson are relevant to daily life	.69	4
Concurrent Learning and Assessment	Assessments help me learn better	.84	3
Bringing Different Viewpoints	Class help me learn that everyone has different view points	.83	5

Correlations between the ACLE scales were computed as to see if they assessed distinctively different aspects of the learning environment. Table 3 represents the inter-correlation amongst the six dimensions in the ACLE.

Table 3. Interscale correlation for ACLE

Student-centered	1.00					
Thought-Provoking	.56	1.00				
Collaborative	.27	.31	1.00			
Life Relevance	.48	.65	.25	1.00		
Concurrent Learning and Assessment	.35	.45	.22	.47	1.00	
Bringing Different Viewpoints	.60	.65	.26	.62	.44	1.00

Note: All correlations are statistically significant at 0.01 level

c) Data Analysis

In data analysis we followed several steps. First, we computed the average perception of the classroom learning environment scores ranged from 1 to 5 for each student for each scale. Then, to address the aforementioned research questions, we used MANOVAs to identify the influences of school location and class size differences in each classroom learning environment score. MANOVA is appropriate when the research wants to compare two or more independent variables in two or more groups (French, Poulsen, & Yu, 2006). All analyses were done with SPSS 21. Because the sample sizes in the groups (rural vs urban location, and small vs large groups) were unequal, the homogeneity of variance was analyzed through Box's M test in SPSS (Tabachnick & Fidell, 2013). The tests of homogeneity of variance were not significant for school location ($F(21, 2588288) = 1.23, p = .21$) and for class size ($F(21, 2890773) = 1.35, p = .13$).

FINDINGS

The mean and standard deviation value for each variable within groups were presented in Table 4. These descriptive statistics (mean scores and standard deviations) may give an idea about the students' perceptions on the constructivist classroom learning environment. The students' perceptions on the constructivist learning environment were between moderate and high with some variance; the means were close to but above the mid-point. The life relevance dimension had the highest mean value ($M = 3.89$), whereas the collaboration dimension had the lowest mean value ($M = 3.31$).

Table 4. Means and standard deviations for variables in the study by location and class size

	Location		Class size		All (N=1882) Mean (SD)
	Mean (SD)		Mean (SD)		
	Rural	Urban	Small (n≤20)	Large (20<n)	
ACLE					
SC	3,72(0.77)	3,68(0.73)	3,69(0,77)	3,69(0,73)	3,69(0.74)
TP	3.84(0.65)	3,71(0.68)	3,80(0,66)	3,72(0,66)	3,74(0.67)
CO	3,38(0.60)	3,28(0.62)	3,38(0,58)	3,28(0,61)	3,31(0.60)
LR	4.04(0.77)	3.83(0.81)	4,00(0,77)	3,85(0,80)	3,89(0.80)
CLA	3,61(0.60)	3,49(0.62)	3,58(0,61)	3,50(0,62)	3,52(0.61)
BDV	3,91(0.70)	3.80(0.70)	3,90(0,71)	3,81(0,70)	3,83(0.70)

Note: SC: Student-centered, TP: Thought-Provoking, CO: Collaborative, LR: Life Relevance, CLA: Concurrent Learning and Assessment, BDV: Bringing Different Viewpoints.

a) Differences between students' perceptions of constructivist environment in rural/urban areas

MANOVA with SPSS 21 were used to address the first research question. MANOVA results revealed a statistically significant difference in students' perceptions of the constructivist classroom learning environment based on school location, $F(6, 1875) = 6.26, p < .001$; Wilk's $\Lambda = 0.98$. Addition to this, the post ANOVA test results revealed that there were statistical differences on dimensions of thought provoking ($F(1, 1881) = 13.55, p < .001$), collaborative ($F(1, 1881) = 9.73, p = .002$), life relevance ($F(1, 1881) = 27.50, p < .001$), concurrent learning and assessment ($F(1, 1881) = 13.95, p < .001$), and bringing different viewpoint ($F(1, 1881) = 8.82, p = .003$). Yet, there was no statistically significant difference on the dimension of student centered based on school location ($F(1, 1881) = 1.32, p = .25$).

Considering mean scores on Table 4, these results point out that the rural school students perceived their classroom learning environment more constructivist than did the

urban school students. More specifically, the students in rural schools more positively reported their constructivist environment in terms of life relevance ($M= 4.04$ vs $M= 3.83$). Additionally, the students on rural location reported perceiving more positive classroom learning environment on thought provoking collaborative ($M= 3.38$ vs $M= 3.28$), concurrent learning and assessment ($M= 3.61$ vs $M=3.49$), and bringing different viewpoints ($M= 3.91$ vs $M= 3.80$) than did those in urban location. Addition to these results, the students in rural schools reported the highest mean score on life relevance, whereas they did the lowest on the collaboration. Similarly, the students in urban school reported highest mean score on life relevance, whereas the lowest mean score was on the collaboration.

b) Differences between students' perceptions of constructivist learning environment in large/small classroom

To address the second research question, again we used MANOVA with SPSS 21. MANOVA results yielded a statistically significant difference on students' perceptions of the constructivist classroom learning environment across small versus large classes ($F(6, 1875) = 3.83, p = .001$; Wilk's $\Lambda = 0.98$). Addition to this, the post ANOVA test results revealed that there were statistical differences on dimensions of thought provoking ($F(1, 1881) = 4.23, p < .04$), collaborative ($F(1, 1881) = 8.19, p = .004$), life relevance ($F(1, 1881) = 13.23, p < .001$), concurrent learning and assessment ($F(1, 1881) = 5.20, p = .023$), and bringing different viewpoint ($F(1, 1881) = 4.93, p = .026$). The post ANOVA revealed no statistically significant difference on the dimension of student centered based on class size ($F(1, 1881) = 0.50, p = .82$).

Considering mean scores on Table 4, these results point out that the students on small classes perceived their constructivist learning environment more constructivists-based than did those in large classes. More specifically, the students in small classes reported their classroom learning environment more life relevance ($M= 4.00$ vs $M= 3.85$), thought provoking ($M= 3.80$ vs. $M= 3.72$), collaborative ($M= 3.38$ vs $M= 3.28$), concurrent learning and assessment ($M= 3.58$ vs $M=3.50$), and bringing different viewpoints ($M= 3.90$ vs $M= 3.81$) than did those in large classes. Addition to these results, the students in rural schools reported the highest mean score on life relevance, whereas they did the lowest on the collaboration. Again, students in both groups reported highest mean score on life relevance, whereas the lowest mean score was on the collaboration.

DISCUSSION and CONCLUSION

The main purpose of this study was to investigate the middle school students' perception of their classroom learning environment by using the ACLE questionnaire and to see if it differed in terms of class size and school location. With respect to the first research question, a statistically significant difference between students' perceptions of the classroom learning environment in small and large classes was found. Students who were in small classes rated their learning environment more constructivist based than those who were in large classes. Students in the sampled small class rated their classroom learning environment the highest on life relevance dimension. They also rated the collaboration dimension as the lowest feature of their constructivist classroom learning environment. These findings supported the previous study have reported that class size is an important factor that effects student' perceptions. Compared to the previous studies conducted in other countries, although Den Brok et al. (2006) found students in large class perceived their constructivist environment higher than those who were in small classes, consistent with Levy et al. (2003), it was found that students in small class was perceived their constructivist learning environment more positively than those who in large class. This may be because of the cultural difference or

other factors dynamic such as age, gender, and teachers and the like, that effect students' perception. Overall, small class enables students to be paid more individualized attentions by their teacher. It may increase student achievement to reduce class size below 20. The issue of class size is essential to increasing our understanding of students learning needs.

As for the second research question, significant difference was found between students' constructivist perception in rural and urban areas. An interesting finding of the study was that students schooling in rural areas perceived their learning environment more positively than those in urban area. Similar to finding in the class size, students in rural and urban area rated their classroom the most constructivist-oriented in life relevance scale and the least in collaboration. Compared to previous studies conducted in other countries, although Young (1998) reported that students in urban areas performed better than those in rural areas, in this study the findings showed that students who were in rural areas rated higher their constructivist environment than those who were in urban areas.

The findings of this study can have practical implication for science teachers. One important finding of this study is that students in large/small classes and urban/rural schools in the sample perceived their constructivist environment less collaborative. Collaboration is an important tool in the knowledge construction and socialization (Arkun & Askar, 2010). Classroom should be the place where students reflect, share and discuss their ideas with each other. Therefore, teachers should pay more attention on how students can collaborate with each other and arrange their classroom practices. Moreover, students should be encouraged to work together by their teachers rather than studying individual. Teachers may design their classroom in a way that students may easily collaborate with each other such as the laboratory design in squared tables, rather than a conference design in their classrooms. Such a learning environment may increase learner interaction, and so provide rich experiences and better learning (Pratt, 2003).

Limitation of study

This study has certain limitations. First, during the validation of the instrument, two items were removed from the analysis. Therefore, further research is needed to validate the questionnaire and duplicate the present study. Second, because we conducted this investigation in only 20 schools in Trabzon, located in Black Sea region in Turkey, future researchers should approach with this caution on generalization of these findings. Third, this study took into consideration two factors, class size and school location, which may affect students' perception. Therefore, the effects of the other factors that literature suggests should be considered.

REFERENCES

- AAAS. (1993). American Association for the Advancement of Science, Project 2061: Benchmarks for Science Literacy. New York: Oxford University Press.
- Abd-El-Khalick, F., & Lederman, N.G. (2000). Improving science teachers' conceptions of nature of science: a critical review of the literature. *International Journal of Science Education*, 22(7), 665-701.
- Akerson, V. L., Buzzelli, C. A., & Donnelly, L. A. (2008). Early childhood teachers' views of nature of science: The influence of intellectual levels, cultural values, and explicit reflective teaching. *Journal of Research in Science Teaching*, 45, 748-770
- Bady, R. A (1979). Students' understanding of the logic of hypothesis testing. *Journal of Research in Science Teaching*, 16, 61-65.

- Bell, R.L., & Matkins, J.J. (2003). Learning about the nature of science in an elementary science methods course: content vs. context. Annual Meeting of the National Association for Research in Science Teaching (NARST), Philadelphia, Pa.
- Boujaoude, S. (1996). Epistemology and Sociology of Science According to Lebanese Educators and Students, Annual Meeting of the National Association for Research in Science Teaching, St.Louis, Mo.
- Brickhouse, N.W., Dagher, Z.R., Letts, W.J., & Shipman, H.L. (2000). Diversity of students' views about evidence, theory, and the interface between science and religion in an astronomy course. *Journal of Research in Science Teaching*, 37(4), 340-362.
- Çelik, S., & Bayrakçeken, S. (2004). Öğretmen adaylarının bilim anlayışları ve “fen, teknoloji ve toplum” dersinin bu anlayışlara etkisi. VI. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi, Marmara Üniversitesi, İstanbul.
- Clough, M.P., & Olson, J.K. (2001). Structure of a course promoting contextualized and decontextualized nature of science instruction. Annual Meeting of the Association for the Education of Teachers, St.Louis, MO.
- Clough, M.P. (2003). Explicit but insufficient: additional considerations for successful nos instruction. Annual Meeting of the Association for the Education of Teachers, St.Louis, MO.
- Demirdöğen, B., Hanuscin, D.L., Uzuntiryaki-Kondakci, E. & Köseoğlu, F. (2015). Development of preservice chemistry teachers' pedagogical content knowledge for nature of science. *Research in Science Education*, doi: 10.1007/s11165-015-9472-z
- Deng, F., Chai, C.S., Tsai, C., Lin, T. (2014). Assessing South China (Guangzhou) high school students' views on nature of science: a validation study. *Science & Education*, 23, 843-863
- Driver, R., Leach, J., Millar, R., & Scott, P. (1996). Young People's Images of Science. Buckingham, UK: Open University Press.
- Govett, A.L. (2001). *Teacher's Conception of the Nature of Science: Analyzing the Impact of A Teacher Enhancement Program in Changing Attitudes And Perceptions of Science And Scientific Research* (Unpublished Phd Thesis). College of Human Resources And Education, West Virginia University.
- Hammer, D. (1994). Epistemological beliefs in introductory physics. *Cognition and Instruction*, 12, 151-183.
- Hogan, K. (1999). Relating students' personal frameworks for science curriculum. *Science Education*, 72, 19-40.
- Jungwirth, E. (1970). An evaluation of attained development of the intellectual skills needed for understanding of the nature of scientific inquiry by BSCS pupils in Israel. *Journal of Research in Science Teaching*, 7, 141-151.
- Khishfe, R., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching*, 39(7), 551-578.
- Khishfe, R., & Lederman, N. (2003). The development of students' conceptions of nature of science. Annual Meeting of the American Educational Research Association (AERA), Chicago, IL.
- Khishfe, R.F. (2004). *Relationship between students' understandings of nature of science and instructional context*. (Unpublished Phd thesis). Graduate College of The Illinois Institute of Technology. Chicago, Illinois.
- Khishfe, R.F. (2015a). Relationship between nature of science understandings and argumentation skills: A role for counterargument and contextual factors. *Journal of Research in Science Teaching*, 49(4), 489-514

- Khishfe, R.F. (2015b). A look into students' retention of acquired nature of science understandings. *International Journal of Science Education*, 37 (10), 1639-1667
- Kılıç, K., Sungur, S., Çakıroğlu, J., & Tekkaya, C. (2005). Ninth grade students' understanding of the nature of scientific knowledge. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 28, 127-133.
- Lecompte, M. D., & Preissle, J. (1993). *Ethnography and Qualitative Design in Educational Research*. (2nd Ed). San Diego: Academic Press.
- Lederman, N.G. (1992). Students' and teachers' conceptions of the nature of science: a review of the research. *Journal of Research in Science Teaching*, 29(4), 331-359.
- Lederman, N.G. (1999). Teachers' understanding of the nature of science and classroom practice: factors that facilitate or impede the relationship. *Journal of Research in Science Teaching*, 36(8), 916-929.
- Macaroglu, E., Taşar, M. F., & Cataloglu, E. (1998). Turkish preservice elementary school teachers' beliefs about the nature of science. Annual Meeting of National Association for Research in Science Teaching (NARST), San Diego, CA.
- Mccomas, W. (1996). Ten myths of science: reexamining what we think we know about the nature of science. *School Science and Mathematics*, 96, 10-16.
- Mcnabb, D.E. (2002). *Research Methods in Public Administration and Nonprofit Management: Quantitative and Qualitative Approaches*. M.E. Sharpe, Armonk, Newyork.
- MEB. (2005). *İlköğretim Fen Ve Teknoloji Dersi Öğretim Programı*. Ankara: Milli Eğitim Bakanlığı Yayınları.
- Meichtry, Y.J. (1992). Influencing student understanding of the nature of science: data from a case of curriculum development. *Journal of Research in Science Teaching*, 29(4), 389-407.
- Moss, D.M., Abrams, E.D., & Kull, J.R. (1998). Describing students conceptions of the nature of science over an entire school years. Annual Meeting of the National Association for Research in Science Teaching. San Diego, CA.
- Moss, D.M., Abrams, E.D., & Robb, J. (2001). Examining student conception of the nature of science. *International Journal of Science Education*, 23(8), 771-790.
- Murcia, K., & Schibeci, R. (1999). Primary student teachers' conceptions of the nature of science. *International Journal of Science Education*, 21(11), 1123-1140.
- NRC (1996). *National Research Council, National Science Education Standards*. Washington, DC: National Academic Press.
- Oyman, N.Y. (2002). *İlköğretim Fen Bilgisi Öğretmenlerinin Bilimin Doğası Hakkındaki Anlayışlarının Tespiti*. (Yayınlanmamış Yüksek Lisans Tezi). Marmara Üniversitesi, İstanbul.
- Peters, E. E. (2012). Developing content knowledge in students through explicit teaching of the nature of science: Influences of goal setting and self-monitoring. *Science & Education*, 21, 881-898
- Roth, W.M., & Roychoudhury, A. (2003). Physics students' epistemologies and views about knowing and learning. *Journal of Research in Science Teaching*, 40, 114-139.
- Rubba, P., Horner, J. K., & Smith, J. M. (1981). A study of two misconceptions about the nature of science among junior high school students. *School Science and Mathematics*, 81, 221-226.
- Rudge, D. W., & Howe, E. M. (2009). An explicit and reflective approach to the use of history to promote understanding of the nature of science. *Science & Education*, 18, 561-580.
- Ryan, A. G., & Aikenhead, G.S. (1992). Students' preconceptions about the epistemology of science. *Science Education*, 76, 559-580.

- Sandoval, W.A., & Morrison, K. (2003). High school students' ideas about theories and theory change after a biological inquiry unit. *Journal of Research in Science Teaching*, 40(4), 369-392.
- Scharmman, L.C. (1988a). Locus of control: a discriminator of the ability to foster an understanding of the nature of science among preservice elementary teachers. *Science Education*, 72, 453-465.
- Scharmman, L. C. (1988b). The influence of sequenced instructional strategy and locus of control on preservice elementary teachers' understanding of the nature of science. *Journal of Research in Science Teaching*, 25, 589-604.
- Smith, C. L., Maclin, D., Houghton, C., & Hennessey, M.G. (2000). Sixth-grade students' epistemologies of science: the impact of school science experiences on epistemological development. *Cognition and Instruction*, 18, 349-422.
- Songer, N.B., & Linn, M.C. (1991). How do students' views of science influence knowledge integration?. *Journal of Research in Science Teaching*, 28, 761-784.
- Strauss, A., & Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures And Techniques*. Ondon: Sage Publications.
- Taşar, M.F. (2003). Teaching history and the nature of science in science teacher education programs. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 1(13), 30-42.
- Tsai, C.-C. (1999). The progression toward constructivist epistemological views of science: a case study of the STS instruction of Taiwanese high school female students. *International Journal of Science Education*, 21(11), 1201-1222.
- Wahbeh, N. & Abd-El-Khalick, F. (2014). Revisiting the translation of nature of science understandings into instructional practice: Teachers' nature of science pedagogical content knowledge. *International Journal of Science Education*, 36(1), 425-466.
- Yakmacı-Güzel, B. (2000). Fen alanı (biyoloji, kimya ve fizik) öğretmenlerinin bilimsel okuryazarlığın bir boyutu olan "bilimin doğası" hakkındaki görüşleriyle ilgili bir tarama çalışması. IV. Fen Bilimleri Ve Matematik Eğitimi Kongresi, Hacettepe Üniversitesi, Ankara.

Effectiveness of the Proposed Training Formative Assessment Programme and its Impact on Teaching Style Improvements of Saudi Science Teachers in Saudi Arabia

Fayadh Hamed ALANAZI¹ 

¹Assistant Professor, Aljouf University, Saudi Arabia

Received: 09.06.2016

Revised: 09.11.2016

Accepted: 01.12.2016

The original language of article is English (v.14, n.1, March 2017, pp.35-56, doi: 10.12973/tused.10189a)

ABSTRACT

This work investigates the effects associated with training initiatives consistent with formative assessment to enhance teaching among science teachers in Saudi primary schools. The programme comprised training sessions that were ongoing for a period of two months, which emphasised the strategic skills of adopting the formative assessment. The researcher adopted an experimental approach centred on the design of the control and experimental groups to establish the overall efficiency of the suggested training programme in the development of knowledge among teachers in the field of formative assessment strategy skills. The sample comprised 33 primary school science teachers who were based in Sakaka in the Al-Jouf region; the sample was randomly divided into an experimental group (16 teachers) and a control group (17 teachers). The findings indicate the success of the proposed program, in which the experimental group was more a function of the control group in the telemetric measurement. Moreover, classroom observations were applied by the researcher, in which the science teachers were assessed with the aim to determine whether a form of formative assessment was used following the completion of the programme. The findings suggest that the science teachers adopted various strategies of formative assessment, including sharing learning goals, open questions, assessment-based evidence, and teacher-child classroom dialogue.

Keywords: Science teachers, summative assessment, formative assessment, training program, teaching.

INTRODUCTION

Without question, assessments play a fundamental role in the enhancement of student learning. As stated by the Task Group on Assessment and Testing Report (TGAT, 1988), “assessment is at the heart of the process of promoting children’s learning” (para. 3). In the specific field of education, there are two recognised forms of evaluations: assessment for learning and assessment of learning.

It is essential to differentiate between assessment for learning, which is recognised as a formative assessment, and the assessment of learning, which is recognised as a summative assessment, to facilitate teachers in their application in practice. As stated by Sadler (1989), the formative assessment is centred on how quality-related judgements that pertain to student



responses may be used to enhance and generally shape the competence of students through short-circuiting the inefficiency and randomness associated with trial-and-error learning. Moreover, the view has also been adopted that a summative assessment is focused on summing or otherwise providing a summary concerning the achievement status of a student, and it is focused on providing a report upon course conclusion, particularly in cases of certification (ibid). Nevertheless, summative and formative assessments do not always lack in compatibility, with the information garnered from summative assessments formatively used when learners apply it to guide their activities and efforts in other courses. Moreover, a formative assessment may be applicable for use in informing a summative assessment, with these assessments indicating whether the learner is ready to progress to further material. After the instructor has established that the learner has adequate knowledge and skills, a summative assessment may be implemented, which enables them to showcase their skills and knowledge.

Furthermore, through the Qualification and Curriculum Authority (2009) in the UK, it has been established that assessment for learning goes against learning assessment, which is recognised as comprising judgements concerning student performance consistent with national standards. It is common for teachers to make these statements following the completion of a work unit, year or key stage. Furthermore, student performances in regards to levels are also described through test results.

From the researcher's perspective, the key difference between the two approaches to assessment is related to the target or objective rather than the time. When the key emphasis is on encouraging the learning of the student, which subsequently forms an efficient aspect of the learning process through the adoption of many different instruments, such as sharing learning goals, self and peer assessments, and becoming involved in discussion, learning evaluation would occur. When the focus of an evaluation is centred on reporting the mid-term progress of a student, for example, this approach is considered an assessment of learning.

Literature Review

The History of Formative Assessment

The term 'formative' was first introduced in the arena of mainstream teaching in the UK when the following statement was made by the TGAT (1988) concerning what should be afforded priority in a National Curriculum: "the results [of assessment] should provide a basis for decisions about pupils' further learning needs: they should be formative" (Paragraph 5, emphasis in original). Importantly, formative assessments were used by the TGAT with respect to planning the subsequent stages, with the theory adopted in support of the view that formative assessments are recognised as pertaining to the application of assessment information by teachers to adapt their planning and teaching; this has been widely recognised in literature that followed in the field of the UK National Curriculum's application. Nevertheless, to make assessment-based decisions, there is a need for the teacher to be well positioned to interpret the assessment outcomes in the correct manner (Frowe, 2005).

In the TGAT model, there is an intention for teachers to utilise the feedback in an effort to make programmatic decisions. Nevertheless, as stated by Sadler (1989), students should be provided with feedback in an effort to identify their strengths and weaknesses and thus recognise the involvement of the student through the establishment of the conditions that may achieve improvement in student learning. The view was also held by Sadler that students should adopt an active role in the participation of self-assessment to ensure that they may achieve the greatest possible formative value from the assessment; moreover, the individual must implement the most suitable actions that may result in improvements, with learners encouraged to select the most suitable strategy or approach to improve their performance.

The perspective adopted by Sadler is comparable to the views of other researchers: for example, the work of Harlen & James (1996) emphasises that formative assessments should enable students to self-assess their progress, as well as enable teaching staff to examine and investigate the areas of strengths and weaknesses in students and subsequently make changes to their learning activities to meet students' learning requirements. Moreover, the work of Black & Wiliam (1998a) further emphasises that "feedback is a concept that is central to formative assessment" (pp. 9–10).

Importantly, Black & Wiliam (1998a) have also made reference to various points concerned with emphasising formative assessment. Student feedback should be centred on the specific qualities of the work, with suggestions made on how the work may be suggested, while ensuring comparisons are avoided. However, from the researcher's own viewpoint, comparison may be positive between students, particularly with good students who require this comparison to feel satisfied with their own efforts. It has also been established by Black and his colleague that pupils would benefit from being trained in self-assessment to facilitate their overall insight into, and understanding of, the main purpose of their work, with additional opportunities regarding how their understanding may be expressed. It is considered that this could be one aspect of teaching.

The Qualifications and Curriculum Authority (2003) has indicated that assessment for learning requires that learning be improved through the use of classroom assessment. Furthermore, an assessment that prioritises serving the purpose of promoting student learning represents assessment for learning (Black at al., 2004). It therefore demonstrates key differences between assessments that are primarily designed with the aim to satisfy accountability or certify competence or ranking.

In conclusion, throughout the earlier stages, formative assessment was applied (1988) with a focus on informing the steps to be taken by the teacher; however, priority is subsequently centred on not only planning teaching but assessment for learning as a principle of teaching and learning, with substantial consideration of participation centred on both teachers and students, such as through sharing learning goals, performing self-assessments, questioning, achieving feedback, completing peer assessments and applying the formative use of summative tests. Notably, however, one of the most in-depth and wide-ranging descriptions of formative assessment was introduced by Shepard (2006), who focused on the works completed by Black & Wiliam (1998b) and Sadler (1989) and devised a framework of formative assessment as an approach to guide student learning, as well as provide information to be used by teachers to enhance their overall instructional practice

It is valuable to recognise that in the history of formative assessment, in 2008, the Council of Chief State School Officers (CCSSO) made formative assessment a critical element of the definition implemented; thus, formative assessment was recognised as a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve student achievements of intended instructional outcomes (Melmer et al., 2008).

In more recent times, Wiliam (2011) has emphasised the value of formative assessment as a critical element centred on the enhancement of overall teacher quality, while also having the most significant effect on student outcomes. He considers several of the more commonly implemented initiatives that seek to enhance student achievement, including learning styles, and presents research that details formative assessment practices as exerting a far more profound effect on educational attainment compared with other reforms.

Effectiveness of Assessment for Learning (AfL) in Enhancement of Student Learning

The achievements of students have significantly improved since assessment for learning practices have become positioned as a fundamental aspect of classroom routines. Evidence

obtained on a global scale indicates that an effect size of half to one, as well as a half standard deviation and more may be directly attributable to the implementation of classroom assessment for learning. Thus, throughout the course of this paper, the effect size is a standardised method of drawing a contrast between the results of multiple pedagogical experiments.

Throughout the mastery learning study presented by Bloom (1984), a summary of the effects into the mastery learning models was provided, which drew a contrast between standard whole-class instruction and two experimental interventions: primarily within mastery learning conditions and secondarily within individual tutorial conditions. The in-depth, wide-ranging application of classroom assessment for learning has become a key element of instruction in both types of experimental conditions. It was highlighted throughout the course of the analysis that there were differences that ranged from one to two standard deviations in the achievements of students who favoured the assessment for learning experimental conditions. Nevertheless, other evaluations that utilised the same experiments achieved more minor effects, such as in the case of Cohen, Kulik & Kulik (1982), where the effective size was only 0.4 standard deviations.

In an assessment of 250 studies performed worldwide, as presented for the period that spanned 1987–1998, Black & Wiliam (1998a) suggested that teachers placed emphasis on assessment for learning rather than assessment of learning, and this focus created a significant increase in the achievements of learners. Regarding formative assessment, the scholars concluded that the ongoing adoption of classroom formative assessments would result in improvements in student achievements by 0.4 to 0.7 standard deviations, which represents a figure sufficiently adequate to raise the United States to the top five in international rankings. Furthermore, it was established that an improved assessment was useful in assisting low achievers more than other students and would therefore decrease the achievement range while also increasing achievements overall. Nevertheless, it was established that teacher-centred teaching and summative-focused assessments achieve positive outcomes in relation to the performance of learners.

The Iowa Tests of Basic Skills were examined by Meisels et al. (2003) via the assessment of the scores of third- and fourth-grade learners who had been involved in the Work Sampling System (WSS) programme; the aim of this initiative was centred on encouraging a curriculum-embedded performance assessment compared with a demographically comparable contrast group. The findings suggest that “Children in WSS classrooms made greater gains in math than children in the other two groups, although the results were only marginally significant when compared with gains by the matched contrast group. Effect sizes ranged from .75 to 1.5 standard deviations” (p. 2). Furthermore, Rodriguez (2004) has discussed smaller size effects with respect to an assessment of the relationship between the achievements of US math students involved in the TIMSS (Third International Math and Science Study) and assessment practices; it was subsequently concluded that students could be helped by teachers, particularly in the case of low-achievers in math, when they were taught that their own success in learning could be controlled.

In the work conducted by Ruiz-Primo & Furtak (2006), four teachers were involved in a five-day training in the application of formal embedded assessments in science instruction. The way in which these assessments should be used to garner immediate feedback for learners was taught. Student performance was measured with the use of pre-tests, embedded assessments and post-tests. It was established that the intervention had a notable effect on student performance. The findings indicated that fidelity and the overall quality of the application varied among teachers, and, in general, teachers with a greater degree of quality implementation reaped better scores in the post-test. Notably, however, it was indicated that the four-teacher sample size is substantially limited. Achievement gains cannot be attributed

to intervention when considering that other teacher characteristics may have contributed to the various differences identified in the performance of the teachers. Importantly, tasks were performed correctly by students; however, the students were often not able to provide explanations as to why they answered a specific way.

Overall, the OECD (Organisation for Economic Cooperation and Development, 2007) has made the suggestion that the studies conducted by the OECD's Centre for Educational Research and Innovation (CERI) support the previously described effects. The equity of student findings would also be improved through formative assessment, which has been applied by the schools' present data not only in mind of attaining general gains in academic achievements but also particularly high gains for students previously recognised as underachieving. The retention of learning and overall attendance have exhibited improvements, in addition to an overall improvement in the work of learners. Various aspects of the work conducted by Black & Wiliam (1998a) have been called into question in the work of Dunn & Mulvenon (2009), with the latter scholars contending that the limited scientific evidence concerning the effects of formative assessment in education is, at least in some regard, a result of the issue that there is no sound consensus regarding what constitutes formative assessment. Without agreement in this regard, the authors state that the establishment of a well-formed body of research is problematic. The authors further challenge the 1998 work as a result of the belief that it exhibits significant reliance on a study that utilised a sample composed of disabled students, which equated to 83%. The authors further emphasise that higher quality students with a more wide-ranging representation of the student population may have resulted in smaller effect sizes. The authors cited numerous concerns in relation to methodological issues in the article Black & Wiliam (1998a); however, they agree that there is support concerning the application of formative assessment and thus indicate the need for higher quality students in an effort to strengthen the research base.

Moreover, Kingston & Nash (2011) completed a meta-analysis, which established more modest mean effect sizes that notably equated to approximately 20; these findings suggest that the work of Black & Wiliam (1998a) was overly generous when estimating the effects on achievements. Numerous different elements were queried in the work of Briggs & Ruiz-Primo (2012) in relation to the study by Kingston & Nash (2011); however, there is a consensus that a notable degree of uncertainty remains in terms of the effect of formative assessment on achievement. There is increasing momentum underpinning the application of formative assessment practices in classrooms; however, the analyses completed by various scholars, such as Bennett (2011), Briggs & Ruiz-Primo (2012), Dunn & Mulvenon (2009), and Kingston & Nash (2011), provide the collective suggestion that the evidence base concerning the effects of formative assessment on the academic attainments of students has remained somewhat questionable. More specifically, they suggest the adoption of two solutions: more suitably designed studies and studies that analyse various elements that affect the overall efficacy of formative assessment.

Professional Development Related to Formative Assessment

Alkharusi (2011) indicates that previous and recent studies of classroom assessment have consistently expressed a concern regarding the sufficiency of teachers' assessment skills. However, a limited number of studies have been recognised as associated with the link between formative assessment and professional development (Cole, 2010). Meisels et al. (2003) have analysed student performance (for further details, refer to previous Section) as an outcome measure associated with the effects of professional development focused on formative assessment. Standardised test scores were analysed in terms of assessing the effects of a three-year implementation of an embedded performance assessment system in schools

self-selected for the programme's application. This study further established significantly increased standardised reading scores for the treatment group.

Furthermore, the work of Wiliam et al. (2004) has detailed numerous different elements inherent in an 18-month professional development programme aimed at providing teachers with support in further improving their application of formative assessment. Twenty-four teachers at the secondary school level (12 mathematics and 12 science) spent 6 months learning about four strategies of formative assessment, namely, descriptive feedback, questioning, self and peer assessment, and sharing learning criteria with students. Action plans were subsequently created by the teachers, which outlined the various aspects of formative assessment to be adopted in their classrooms. Throughout the duration of the school year, teacher observations were performed, with the opportunity to consider and discuss with other professionals how the formative assessment practices detailed in their action plans could be implemented. The findings suggest that individuals involved in the development demonstrate notable changes in their perceptions of themselves as professionals.

A study that utilised a sample of 25 secondary teachers in math and science, which was conducted in New South Wales Australia by Panizzon & Pegg (2007), aimed to investigate assessment and its changing nature, as well as the effects on in-classroom practices. Across a two-year study, teachers attended numerous professional development sessions and received continuous consultative support. All sessions were recorded and subsequently transcribed; interviews were completed with all teachers at the end of the school year. The key aspect inherent in the study was the questioning, whereas the six contributing aspects included assessment linked to pedagogy, attention to cognition, classroom advantages for students, classroom advantages for teachers, pedagogical practices and teaching strategies. These findings signify a significant change in the perceptions of teachers regarding assessment, which redirects the focus from the accumulation of student marks to diagnosis with the objective to direct teaching towards improving the mathematical and scientific understanding of students.

Teachers were provided with a short-term professional development session by Yin et al. (2008) with the aim to teach them how a particular embedded assessment could be applied in their science classroom. The scholars designed and implemented formative assessments in an inquiry science-based unit. The sample utilised comprised 12 middle-school science teachers and their students, all of whom were randomly selected and subsequently assigned to control or experimental groups, with 6 students per group. The embedded assessments were provided to the control teachers, without the inclusion of training. The findings indicated that teachers significantly varied in terms of their effects on achievement, conceptual change and student motivation. Nevertheless, the effects of the formative assessment treatment on these outcomes were not significant.

In a study by Phelan et al. (2009), 91 teachers were randomly assigned within schools to treatment or control groups. For the treatment, teachers were provided with training on three lessons in the 6th grade, which were to be taught throughout the year, with an emphasis placed on math and formative assessments. The POWERSOURCE© intervention specifically directs attention to middle-school-level mathematics, notably beginning at the 6th grade, as well as aiding in assuring that learners have the understanding required for them to be successful in the algebraic field. This particular intervention focuses on big ideas and associated skills in four specific fields that underlie algebraic success: (a) rational number equivalence (RNE), (b) properties of arithmetic (PA—the distributive property), (c) principles for solving linear equations (SE), and (d) the application of core principles in these areas to other critical areas of mathematics, namely, geometry and probability (RA). These areas were selected because of their perceived value in subsequently mastering algebra, as well as with respect to their valuable position in state mathematics standards across the 6th grade.

In contrast, the individuals in the control group were not provided with training; instead, they were instructed to teach the students the same math topics with their own lesson plans. Furthermore, following each of the three lessons, the teachers in the experimental group met with the researchers to review the work of the learners and subsequently analyse common errors. In a comparison of the conditions, when considering student performance of standardised math outcome measures, as well as teacher math outcomes, the findings indicated numerous key differences in favour of the treatment group. For example, overall, the POWERSOURCE© students did not outperform the students in the control groups, when considering that the researchers did not identify statistically significant main effects associated with the treatment. Notably, however, they identified a notable interaction between 39 treatment and pre-test scores, which implies that the learners with higher pre-test scores are more likely to draw benefit from the intervention formative assessment when contrasted with the learners with lower pre-test scores. The student effect size is as much as a 0.5 pooled standard deviation; thus, the intervention was more significant in relation to the higher performing students compared with the lower performing students. The PA items were recognised as being more problematic for students compared with the items concerned with other domains. Moreover, qualitative data obtained through the completion of the teacher observations and interviews highlighted that teachers recognised their students as experiencing greater problems with the PA content compared with the other POWERSOURCE© areas. Thus, there was a greater impact of POWERSOURCE© on the more difficult items.

The Teacher's Role in Formative Assessment

The adoption of formative assessments is common when teachers seek to change their teaching methods in line with student learning or when they seek to enable students to make changes to their learning in alignment with their performance feedback (Gibson & Shaw, 2010). Formative assessment is diagnostic in nature because it provides insight concerning student learning that may subsequently be applied to change and enhance both learning and teachers to improve the fulfilment of student needs (Black & Wiliam 1998b). Making changes to assessment-centred practices in schools necessitates more in-depth, precise conceptual changes among educators, as well as critical changes in thinking with respect to teaching and assessment and the overall link between the two factors (Earl, 2006).

The implementation of formative assessment in a practical situation further necessitates that teachers make changes regarding how they work with their students. It is not only necessary to enhance existing practices but also requires practice to be reconsidered to improve the essential feedback loops involved, which assist students in interacting and collaborating more in their overall learning and assessment. These changes in assessment are centred on the views that all students can learn and the role of the teacher is to establish numerous ways to ensure that this particular learning occurs (Earl, 2006). Assessment essentially becomes a fundamental element of the learning–teaching process. Thus, formative assessments that are conducted throughout the learning process itself enables changes in teaching and learning (Gibson & Shaw, 2011). This is a novel approach to assessment for the majority of teachers, and this method is in contrast to more long-term experiences and practices of teachers (Earl, 2006). Studies have emphasised that the way in which a teacher adopts formative assessment within the classroom environment ultimately depends on the teacher and their individual experiences (Stephens, 2011).

In providing effective formative assessments, dialogue was highlighted as paramount (Fox-Turnbull, 2006). More specifically, in the current work, it has been established that providing students with the opportunity to discuss their work not only aids them in their knowledge and growth but also provides teachers with insights into the way in which their

learners think. Thus, teachers may utilise this as an opportunity to respond to students in relation to the practicality, usefulness and quality of their thinking, such as through posing questions in an effort to expand understanding (Fox-Turnbull, 2006). Nevertheless, there is also a need for teachers to enable students to have the opportunity to discuss their thoughts in rationalising and explaining their choices (Fox-Turnbull, 2006). One important finding in the work of Fox-Turnbull (2006) is that a significant factor is teacher knowledge, which has a significant impact on the quality of the feedback provided to students by their teachers.

Buck & Trauth-Nare (2009) have suggested that few teachers have a real understanding regarding the pedagogical implications of scaffolding or their role in the application of formative assessments. Thus, Buck & Trauth-Nare (2009) conducted a study centred on the development of understanding the experience of being a teacher, which aims to improve their levels of knowledge through formative assessment and accordingly utilise their understanding with respect to enhancing teacher education practices. The further observation was made that questioning the class was an approach widely applied by teachers; nevertheless, most questions analysed only what students could remember. The findings of the work draw the conclusion that following the implementation of formative assessment methods within the classroom, teachers soon began to recognise an increase in student participation in the development of formative assessment. Moreover, in contrast to high-achieving students, other learners who had functioned unsuccessfully accepted the formative process and began to thrive in the classroom.

The ways in which science teachers have analysed their roles in the assessment of students and teaching have been investigated by Gioka (2009), in which the emphasis of the study was centred on teacher assessment and the feedback provided to learners. It was considered that teachers face numerous challenges; teachers are required to instruct and further support the learning of students. In contrast, teachers must also assess and analyse and subsequently grade students (Gioka, 2009). Overall, teachers provided limited feedback and students were left to make decisions regarding how they could progress their learning.

How teachers apply formative assessments was examined by Stephens (2011) in the context of middle-school classrooms. Data were obtained through the completion of field observation, teacher interviews, and written responses to student work; the results demonstrated that the adoption of formative assessment creates structure and subsequently achieves student engagement. Positive communication among students is pivotal not only in improving the levels of student motivation but also increasing their willingness to participate and cooperate. Providing learners with assignments or activities means there is less discussion time. Moreover, providing learners with the opportunity to work in small groups may also provide a platform for students to learn more through questions. The findings further emphasise that the lack of formative assessment may induce resistance to what the teacher aims to achieve.

Formative assessment in Saudi Arabia:

From a historical perspective, summative assessments, which emphasise the generation of results that highlight student performance, have been demonstrated to be the key assessment practice in schools in the Kingdom of Saudi Arabia (KSA): the examination system is recognised as being the only educational assessment tool in adoption (Al-Sadaawi, 2007). However, in recent times, Saudi Arabia has implemented a shift towards a more constructivist approach in the educational domain, which highlights problem-solving (Alaudan, 2015), analysis and research as opposed to mere repetition and memorisation (ibid). Effective assessment approaches based on constructivist views promote the integration of assessment and teaching (Bekiroglu, 2007). However, regardless of these changes, there has been the neglect to consider FA, which is regarded as best utilised in a constructivist

environment (ibid). This study poses the view that the reason for the failure to adopt formative assessment is that education in the KSA is associated with a lack of teaching training with respect to implementing an efficient assessment.

Professional development is rarely provided to science teachers in the KSA (Al-Sadaawi, 2007), with the suggestion made by Doran et al. (1994) that teachers have limited training in teacher education programmes regarding how classroom formative assessments should be conducted. Moreover, they also have limited technical assistance across their day-to-day professional practices. In addition, it is stated by Berry (2008) that numerous teachers are not even provided with adequate professional development on learning assessments. In this area, it was also established that Saudi teachers lack the necessary skills to evaluate students in the most suitable approach.

Consistent with these findings, the present work poses the argument that there is a pressing need for teaching staff to acknowledge the aim and overall method of formative assessment, as well as the effective tools of formative assessment, including collaboration among students and teachers, explicit learning goals, descriptive feedback to students with regards to their performance, and self and peer assessments. In this regard, the view is adopted by Yin (2005) that teachers engaged in training programmes should be provided with formative assessment tools and helped in their application.

The present work considers that there is a need for relevant professional training to be delivered to teachers to ensure the assessment results for improving education in schools may be used. It is held that student learning may be enhanced through formative assessment, as established in other works previously discussed, which indicates professional development on formative assessment should be focused on improving student attainment. Thus, there is a need for formative assessment to be one element of professional development because this is recognised as a fundamental aspect of teaching, which encourages the higher-order thinking of students and further assists students in satisfying the standards to a high level of proficiency (Trumbull & Gerzon, 2013).

METHODS

a) Research Methodology

Both quantitative and qualitative methods were used in this study.

b) Research Question

What are the effects of the proposed training programme based on formative assessment to improve instructional practices among Saudi science teachers in primary schools?

c) Quantitative Approach

The researcher adopted the experimental method based on the design of the control and experimental groups to determine the effectiveness of the proposed training programme in the development of strategies for using formative assessments among Saudi science teachers in primary schools. The researcher designed the programme. The reason for using the current experimental approach was because the researcher controlled the independent variable, namely, the proposed training programme.

c1) Sample

The study population comprised all science teachers in primary schools in one educational administration in Saudi Arabia (Aljouf); the total number of teachers in this administration was 120. The specific administrations were selected because it would be easier to conduct the study where the researcher lived and worked. The administration department of the Saudi Ministry of Education was asked for a list of e-mails for all science teachers in Aljouf City where the research was conducted. An email was subsequently distributed among all teachers (120). After four weeks, 33 teachers had responded. The researcher subsequently sent another e-mail to the 33 teachers to explain the purpose of the current study and the other processes of investigation (knowledge test, the proposed training programme based on formative assessment).

Sixteen teachers agreed to participate in the experimental group. Thus, the study sample consisted of 33 science teachers from primary schools in Sakaka in the Al-Jouf region, who were divided and randomly assigned to an experimental group (16 teachers) and a control group (17). As a result of the influence of the Saudi culture and considerations regarding religion, all schools are single-sex in Saudi Arabia. The gender separation also indicated it was obligatory to conduct the study from boy only schools and among a sample of male teachers.

From a research perspective, the practicalities of the Saudi context indicated that the study could only include a sample of boys. It is expected that differences may exist between boys and girls as a result of the ways in which different genders are socialised in Saudi Arabia and considering that the inclusion of girls as participants may help establish new and important issues in Saudi education that otherwise may not be identified with boys alone. Furthermore, it should be noted that the average experience is very tight between teachers, i.e., an average of ten years. However, the homogeneity of the two samples was ascertained before applying the proposed training programme through the use of the test T (refer to the table below).

This type of test demonstrated that there are no statistically significant differences between the experimental and control groups in terms of the total score on the questionnaire for testing the level of knowledge in relation to the strategic skills of formative assessment; this finding indicated the homogeneity of the two groups prior to the application of the proposed training programme.

Table 1: Differences between the averages (the level of knowledge in relation to the strategic skills of formative assessment) of the experimental and control groups prior to the application of the proposed training programme.

Statistics group	N	Mean	Std. deviation	df	T	Sig.
Exper	16	168.94	17.510	31	0.772	0.446
Cntrl	17	173.65	17.511			

Table 1 indicates that there are no statistically significant differences between the averages of the experimental and control groups in terms of the degree of knowledge test to measure the level of science teachers on the use of formative assessment strategies; these findings indicate the homogeneity of the two samples.

c2) Tools of the Quantitative Approach

The tools used in this approach were as follows: first, a knowledge test was prepared by the researcher to measure the effectiveness of the proposed training programme in developing the strategies and skills of formative assessment use among Saudi science teachers

at the primary level; second, a training programme was implemented (refer to the following section), which encompassed 15 training sessions for two months, with only the experimental group exposed to this programme. The test comprised 20 questions, which were multiple-choice in nature. The test questions were devised by the researchers and were prepared consistent with a literature review. A pilot that utilised a sample of 5 science teachers was conducted. With the use of the KR-20 (Kuder-Richardson), the test reliability was established. The result was identified as 0.85. The test validity was determined and validated by five experienced science teachers and six professors. As implied by this study, knowledge improvements may result in improvements in instructional practices. Importantly, the knowledge test used the Arabic language.

c3) Programme Session Description

The programme used in the current research was a 15-training session focused on the strategic skills of formative assessment use among Saudi science teachers at the primary school level. The training sessions revolved around the following themes:

- The differences between behaviourist learning theory and constructivist learning theory.
- The differences between formative assessment and summative assessment.
- The differences between learning and teaching, as well as assessment and evaluation.
- Training the teachers in the use of formative assessment strategies: for example, explicit learning goals, collaboration among teachers and students, self and peer assessments, and descriptive feedback to students regarding their performance.
- Provide teachers with practical examples of the importance of formative assessment in improving teaching.

c4) Statistical Analysis

This study aimed to identify the effects of the proposed training programme on the experimental group compared with the control group; thus, comparisons between the experimental and control groups were conducted, as well as between the tribal and dimensional measurements using T-tests to denote differences between the two groups via SPSS (Ver programme. 20.0).

d) Qualitative Approach

The researcher implemented a classroom observation with science teachers who were subjected to the proposed training programme. The reason for this approach was the need to determine the effectiveness of the proposed training programme in improving Saudi teaching in actual classrooms.

d1) Sample

Five science teachers were randomly selected from the experimental group (N = 16). The researcher used classroom observations with the five teachers prior to the proposed training programme to investigate whether the teachers used formative assessment strategies; after the programme, this method was used to determine what strategic skills were used by the teachers in relation to the formative assessment. It should be noted that interviews are not customary in the Saudi context and with this instrument (EL-Deghaidy et al, 2015). However, Flick (1998) indicated that in addition to the competencies of speaking and listening, which are used in the interviews, observation is another everyday skill that is methodologically systemised and applied in qualitative research. It is because the observer may use more than

one sense; he/she may simultaneously smell, touch, hear, and see what is observed (Flick, 1998). In the current research, I observed teachers and the way they benefited from the proposed training programme in relation to the formative assessment in class.

d2) Tools of the Qualitative Approach

Different tools were used in this approach: first, a digital recorder was used to record the session; and second, card notes were used to facilitate the codification of the items that could not be recorded or documented by the recording device. It should be noted that the researchers were non-participant observers; this role was helpful in understanding what occurred from the inside. The data obtained through the use of classroom observation were digitally recorded. Digital recording helped the researcher to capture all teacher and student conversations. Two digital recording devices were used during the entire class observations: one device was placed close to the board where the Saudi teacher frequently stood, and another device was placed at the other end of the room. Each voice recording was downloaded and converted to specific file formats.

Verbatim transcription was implemented because this approach could provide a full transcript of everything that was said in the recording, including the common instances of 'erm' and 'er', 'you know' and 'I mean'. Thus, after completing all transcriptions, the decision was made to send, via email, all files of the digital recordings from the classroom observations to two colleagues who work as Arabic language teachers. They were asked to check the transcripts and review them alongside the digital recordings to provide the percentage of accuracy in the transcriptions. Overall, the feedback was helpful, and the percentage of accuracy was very high, i.e., greater than 99%. All data were collected in Arabic, whereas the research report was written in English. Thus, all data included in the research report had to be translated from Arabic to English. The researcher decided to translate all data into English that could help achieve holistic rather than piecemeal data. Thus, the researcher also decided to use machine translation (MT); this approach is currently used by many professional translators to assist their work (Aiken and Balan, 2011). The original transcription and translation drafts were sent to two Saudi friends of the researcher, who are currently earning their PhDs in Biology at Glasgow University and Science Education at York University. The researcher asked these individuals to follow the same process used when the researcher translated the original transcription. The researcher subsequently asked them to provide a percentage of the accuracy in the work. The average results demonstrated greater than 98.5% accuracy.

d3) Analysis of the Qualitative Approach

A content analysis and conversation analysis were used in this approach. The content analysis is a method that aims to investigate the meaning and content of a text (Hyde et al., 2004). A conversation analysis was applied in an effort to examine the teachers and children's discussion in interaction.

FINDINGS

a) Presentation of Quantitative Findings

This section presents the quantitative findings based on T-tests to determine the effects of the proposed training programme in developing the level of knowledge in relation to the strategies and skills of formative assessment. For example, Table 2 indicates that there is a significant difference between the averages of the tribal measurement before the implementation of the programme and the telemetric after the implementation of the

programme, in which the experimental group was in favour of the dimensional measurement. This finding infers the success of the proposed training programme in the development of the level of knowledge in relation to the strategic skills of formative assessment among Saudi science teachers in primary schools.

Table 2: T-test of the pre and post measurements in the experimental group

Statistics Variable	Sample size	average	Std. deviation	Degrees of freedom	T- value	Significance
Pre	14	169.93	18.43	13	2.66	0.02
Post		185.68	16.11			

Furthermore, Table 3 indicates there was no significant difference in the tribal telemetric measurements in the control group. Thus, the group that underwent the programme was the only group affected; the control group was not affected.

Table 3: T-test of pre and post measurements in the control group

statistics Variable	Sample size	average	Std. deviation	Degrees of freedom	T- value	Significance
Pre	11	173	18.65	10	2.15	0.06
post		159	12.53			

Moreover, Table 4 indicates the success of the proposed programme as the experimental group was more a function of the control group in the telemetric measurement. This finding suggests the effect of the proposed training programme in the development of the level of knowledge in relation to the strategic skills of formative assessment among Saudi science teachers in primary schools.

Table 4: Differences between the experimental and control groups in the telemetric based on a T-test

statistics Variable	Sample size	average	Std. deviation	Degrees of freedom	T- value	Significance
Experimental	14	185.86	16.11	23	4.55	0.001
Control	11	159	12.53			

b) Presentation of Qualitative Findings

Observations were made both before and after the implementation of an intervention (the proposed training programme was based on the strategic skills of formative assessment).

b1) Observations prior to Intervention

During the classroom visit, all teacher/student interactions were captured during the lessons to determine whether Saudi teachers use formative assessment in their classrooms. The results demonstrate that all five Saudi science teachers in primary schools often focused on transferring only scientific information to their pupils. Therefore, pupils are viewed as empty containers, in which teachers deposit information into their minds. Pupils are treated as objects that are acted upon rather than as active and knowledgeable participants in the educational process. Moreover, they tend to use chalk-and-talk methods to explain their lessons, with pupils spending most of their time listening to the teacher and copying from the board, as opposed to being allowed to express their ideas and reflect their attitudes. At the end of the lessons, the teachers allow one pupil to answer when they ask them one or two questions to assess the pupils' understanding. However, if the pupils do not understand, the teachers cannot explain the entire lesson again; therefore, none of the questions appear to seek information unknown to the teacher. This result confirms that the teachers do not use formative assessment.

They also concentrate on improving only good pupils' performances, which requires minimal work, rather than focusing on less-able pupils, who require more time and hard work during lessons. Thus, during the classroom observations, the pupils had a lack of good support and positive feedback from their teachers when they lacked the confidence to be successful in the tasks set by their teachers. Furthermore, Saudi teachers pay close attention to the curriculum and follow textbooks that are extensive and very detailed, particularly in primary schools. Thus, the role of teachers is only concerned with covering the entire lesson, with the teacher often moving from one point to another point without assessing or checking pupil understanding. This is mainly because teachers tend to consider what they will teach rather than what pupils should learn and will understand at the end of lessons.

b2) Observations after Intervention

Five themes may be identified in relation to the formative assessment strategies used by the teachers during the interactions with the pupils, namely, sharing learning goals, open questions, assessment-based evidence, classroom dialogue (teacher-child dialogue) and classroom dialogue (teacher-child dialogue).

Sharing Learning Goals

This transcript indicates that the teacher wanted to share learning goals in terms of thinking and learning, rather than only knowing what they have to do:

Transcript 1/Classroom (5th Grade):

The teacher began to write the following on the board:

'What we will learn today includes:

How does lightning occur?

How does thunder occur?

What is the difference between lightning and thunder?'

T: Today, we will have a very important and interesting lesson, and you can see on the board the objectives that we are going to learn.

T: What will we learn today?

Students: How lightning occurs (choral answer).

T: What else?

S: Why thunder occurs and the difference between lightning and thunder (choral answer).

The teachers used the learning goal to describe what the pupils should know, understand, or be able to do by the end of the lesson. Learning goals are important for children because they are only able to assess their progress when they have a clear notion of what they aim to achieve. However, Clarke (1998) indicates that all too often, children do not have a clear idea of the purpose of their activities; therefore, classroom activities appear to children as a collection of disconnected and meaningless exercises. The key to solving this issue is through pupils sharing learning goals in terms of thinking and learning, rather than only knowing what they have to do. For example: *'What will we learn today?' 'Today, we will have a very important and interesting lesson, and you can see on the board the objectives that we are going to learn.'*

Open Questions

This transcript indicates that the teacher asked open questions to encourage children and provoke a range of answers; this approach provides access to their views regarding things and their feelings, as well as promotes enquiry by the children.

Transcript 2/Classroom (5th Grade):

T: Based on your thinking, why does thunder occur?

S 2: Voices in the sky when the storm comes.

S 5: When the rain falls.

S 7: The sound is caused by static electricity.

T: Who has any other views?

S 13: God wants it to occur.

S 4: It is the air moving.

Thus, the teachers avoid asking their pupils subject-centred questions that focus on only the subject matter; instead, the teachers wanted to consider using person-centred questions, such as *'Based on your thinking, why does thunder occur?'*. The use of the phrase *'Why do you think'* encourages a pupil with an opinion or a pupil who has been thinking about this concept or has ideas regarding it to attempt to answer, regardless of whether they are right or wrong (Harlen & Qualter, 2007). In relation to the question content, Elstgeest (2001) states that productive teachers are more useful in helping children's investigation and thinking. Although the pupil provided the scientific answer, *'The sound is caused by static electricity'*, the teacher attempted to direct the pupils to provide unexpected answers, not merely scientific answers, *'Who has any other views?'*. This approach led the children to provide different types of thoughts, such as religious thinking, *'God wants it to occur'*, as well as everyday thinking, *'It is the air moving'*. This is mainly because the pupils received the message that they are required to conceptualise their own answers.

Assessment-based Evidence

This transcript indicates that the teacher used evidence of the pupils' mastery status to determine whether it was necessary to make adjustments in what they are doing in relation to their instruction.

Transcript 3 /Classroom (4th Grade):

T: Do lightning and thunder occur at the same time?

T: The teacher asked the students to express their opinions through the strategy of thumb, which enabled the teacher to obtain an index survey of the students' understanding; thus, when the student indicates it does not occur at the same time, he puts the finger of a thumbs-up in front of his chest, whereas the students who disagree direct their thumbs to the bottom. This approach enabled the teacher to see the students' answers, whereas the students cannot see each other.

*The majority of the students put their thumbs up.
T: Let us move on to another task.*

The teacher used the thumb strategy to establish immediate feedback in relation to the children's understanding of lightning and thunder. Thus, it is clear from the previous transcript that formative assessment supplied assessment-based evidence as an indication of whether the teachers and students 'got it right'; this may play an important role in helping teachers and students to monitor the effectiveness of their efforts. The majority of the students put a thumbs-up in relation to the question 'Do lightning and thunder occur at the same time?', which indicates there is no need for the teacher to adjust the currently instructional approach. This allows the teacher to know what he planned to do next instructionally: 'Let us move on to another task'. Furthermore, the teachers ask questions for various purposes, such as to initiate a discussion, engage pupils in effective thinking regarding a topic (refer to the previous transcript), close a discussion (the current transcript), or quickly determine whether children follow what has been explained or whether there is confusion.

Assessment methods that assess pupils and provide feedback during teaching are essential. Teachers should gather different information in different forms at different times to make their decisions regarding pupil learning (Bekiroglu, 2007). To improve student learning, teachers must constantly support student participation, assess their understanding and provide regular feedback (Cimer, 2007).

Thus, the need to think deeply about the purposes of posing questions in the classroom should be considered in-depth, with these questions enhanced for use in the assessment of children's performance for learning. This analysis suggests that there is not only one answer in relation to whether closed questions represent a good or bad thing.

Classroom Dialogue (Teacher–Child Dialogue)

This transcript indicates that the teacher allowed every pupil to think, with pupils instructed not to raise their hands if they know the answer to a question.

Transcript 4/Classroom (4th Grade)

T: In this lesson, there is no correct or wrong answer. All answers are very important in enriching the lesson and purpose of the lesson. The most important point is that everyone expresses their opinion and way of thinking. I hope everyone cannot raise their hands; it is necessary for everyone to participate.

Student 3: Lightning comes from God.

Student 4: Lightning comes from God who created everything.

Student 5: Lightning divides the two halves of trees.

Students 6: Lightning comes from the wind.

Students 7: Lightning comes because of the air that charges the water.

The teacher wanted to teach the children to raise their hands at the same time and consider only their thinking, regardless of whether they were correct. 'In this lesson, there is no correct or wrong answer. All answers are very important.' Thus, the teachers created a supportive environment to help pupils to be comfortable with providing a wrong answer (Black et al., 2003), particularly with very young children, who tend to not answer because they are afraid of receiving negative feedback. This approach enabled the children to want to talk about the lightning from different aspects, views and thinking, such as religious thinking ('Lightning comes from God'), everyday thinking ('Lightning divides the two halves of trees') and scientific thinking ('Lightning comes because of the air that charges the water').

Classroom Dialogue (Teacher–Child Dialogue)

This transcript indicates that the teacher initiated a dialogue; groups of pupils were subsequently given time to discuss issues with one another and were invited to report their ideas to the class.

Transcript 5/Classroom (5th Grade):

The teacher divided the students into groups, and each group was provided with a small white paper to write on.

T: Thunder and lightning occur at roughly the same time; however, you see the flash of lightning before you hear the thunder. Why?

The teacher provided the students with three minutes to think and asked a member of each group to speak and provide their answers to encourage each group to discuss with one another.

Group A: It cannot occur. This task is wrong. The lightning and thunder cannot occur together. Lightning occurs first, followed by thunder.

Group B: We can only see the lightning, and we cannot see the thunder.

Group C: Lightning and thunder are voiceless voices.

Group D: Lightning is faster than thunder.

Group A: Light is faster than sound.

Group B: Light comes first; then comes the sound.

Group C: Light is faster than sound.

Group D: Light is faster than sound.

T: What do we conclude from this?

Group D: Light is faster than sound, which indicates the lightning occurs first and then the thunder; however, they occur at the same time.

The teacher used a group discussion strategy because it is recognised as a good approach to practice pupils' cooperation and communication knowledge (think, talk and discuss with one another). The advantage of using this strategy is that pupils who have trouble in calculations or understanding the concept may learn from group members and receive advice from one another. This is clear from Group A, who completely refused the task: *'It cannot occur. This task is wrong. The lightning and thunder cannot occur together. Lightning occurs first and then the thunder.'* However, following the interactions of the pupils in the different groups, such as B, C and D, and without intervention from the teacher, the pupils in Group A modified their knowledge and provided a correct answer: *'light is faster than sound'*. The same situation and process occurred in Groups B and C. The opportunities for the interactions between the groups included an immediate assessment of the pupils' learning and understanding regarding a topic and provided dynamic and live feedback. This work tends to be formatively assessed; thus, the pupils feel able to develop their ideas over time. This is clear from the pupils in Group D with respect to how they developed their answers and justifications for the task provided by the teacher: *'Light is faster than sound, which indicates the lightning occurs first and then the thunder; however, they occur at the same time'*. In terms of assessment, the high level of interactivity in the different groups of teaching (A, B, C and D) indicates that teachers may provide a formative (or ongoing) assessment of progress.

DISCUSSION and IMPLICATIONS

This section discusses and synthesises the findings from a view to address the research question: What are the effects of the proposed training programme based on formative assessment to improve the instruction among Saudi science teachers in primary schools? From the analysis of the quantitative findings, it may be interpreted that there was a statistically significant difference between the experimental and control groups in favour of

the experimental group. Thus, the proposed training programme for two months (60 hours) was successful in developing knowledge related to strategy skills of formative assessment among Saudi science teachers in primary schools. This finding is consistent with Wiliam et al. (2004), in which teachers who participated in professional development based on formative assessment for 18 months significantly changed their perceptions of themselves as professionals.

However, this finding is inconsistent with Yin et al. (2008), who provided science teachers with a short-term professional development session based on formative assessments. The results of the Yin et al. (2008) study demonstrated that the impact of the formative assessment treatment on these outcomes was not statistically significant. Thus, this finding was unanticipated and therefore suggests that teachers who attend a one-off professional development workshop or who are merely provided with new instructional materials make few changes to their practice (Meiers & Beavis, 2005). Darling-Hammond et al. (2009) demonstrated that professional development programmes that provide 30 to 100 hours (an average of 49 hours) over a period of 6 to 12 months had a “positive and significant effect” on student achievement (p. 9). Professional development that provided a limited number of hours (5–14) produced no significant impact on achievement. This finding has important implications for the Ministry of Saudi Education that providing support for science teachers is critical to implement what they are learning.

Teachers should be trained on assessment programmes to be updated with new moves (Alotabi, 2014). However, professional development requires strong, highly visible, organisational support and may be implemented in many ways (Almazroa & Alshamrani, 2015). Sufficient time, resources, and professional assistance should be provided to support Saudi teachers to integrate new knowledge and skills into everyday practices (ibid). Organisational support may represent a key to the success of professional development efforts (ibid) when Saudi teachers participate in a professional development programme on formative assessment, in which teachers must implement their training in schools where students are only graded on a summative assessment (ibid).

Furthermore, the education programmes of Saudi teachers should address higher-order thinking skills in a concrete way to help teachers observe what they look like at different levels and determine how to teach pupils the thinking skills required to complete projects. In addition, teachers need instruction and practice in the skills important for assessment, such as recording anecdotal observations, providing written and oral feedback, and analysing information collected from various types of assessments (The National Foundation for the Improvement of Education (NFIE), 1996).

Furthermore, for the qualitative findings, it is somewhat unexpected that due to the effectiveness of the proposed programme the teachers used different types of formative assessment strategies, such as sharing learning goals, open questions, assessment-based evidence, classroom dialogue (teacher-child dialogue) and classroom dialogue (teacher-child dialogue). These strategies play a crucial role in helping teachers to improve their instruction. For example, prior to the intervention, the Saudi teachers asked very few closed questions to seek the correct answers.

In contrast, after the intervention, the Saudi teachers asked different types of opened and closed questions, and the target was to encourage more students to take turns to think, talk and argue. This interaction helped the Saudi teachers to view the learning as an active rather than passive process and attempt to practice a student-centred approach instead of the teacher-centred approach that is common in Saudi schools. These findings are consistent with Buck and Trauth-Nare (2009) who demonstrated that following the implementation of formative assessment strategies within the classrooms, successful science teachers asked open questions to gain the trust of their students so the students would not shut down because they do not

know the correct answers. The current findings indicate that the teachers used the expression *“there is no correct or wrong answer. All answers are very important in enriching the lesson and purpose of the lesson. The most important point is that everyone expresses their opinion and way of thinking”*. Thus, children are important as individuals, social beings and inquirers. This is crucial to shift the learning from traditional teaching and constructivist learning (ibid).

In another example, a group discussion strategy helped Saudi teachers to assess the students in a way that enabled them to actively engage in making cumulative progress during the process of teaching and learning, rather than viewing the assessment as a product of the learning experience, even when final/modular examinations dictate formal results (Black & Wiliam, 1998b). Formative assessment aims to support student development rather than provide definitive and ‘summative’ measures of progress. A group discussion strategy improves the Saudi teachers’ instructions because they do not emphasise providing a didactic lecture, which covers the subject matter; instead, they adopt the role of a facilitator, who “provides guidelines and creates an environment” for children to arrive at their own conclusions (Alao, et al., 2010, p.56). When formative assessment is conducted during teaching, it provides “both students and teachers with feedback concerning the learning and the teaching of the lesson” (Bulunuz & Bulunuz, 2013, p.130).

This finding runs parallel with Panizzon & Pegg (2007) who indicated that when teachers attended a series of professional development sessions based on formative assessment, they had a major shift in their perceptions of assessment from a focus on the accumulation of student marks to diagnosis as a means of directing teaching to enhance students’ scientific and mathematical understandings.

Moreover, Black & Wiliam (1998a) demonstrated that meaningful formative assessment is not a key priority for many teachers because teachers meet challenges to implement formative assessment in their classrooms and schools. This study begins to highlight several implications for the Ministry of Saudi education, which is ultimately responsible for the educational system and practices. Many of the Saudi curriculum materials commonly in use do not provide the type of curriculum-embedded assessments that evoke pupil thinking and do not provide interpretive frameworks for teachers to use. “There is need to increase the number of formative assessment probes in science textbooks and other resource materials” (Bulunuz & Bulunuz, 2013, p.132).

The implementation of formative assessment may be limited by other conditions in the classroom. Carless et al. (2003) determined that teachers at all levels tend to experience heavy workloads; under these circumstances, it may be a challenge to expect them to engage in learning-centred discussions with individual pupils. When teachers are overloaded with numerous responsibilities, they may perceive formative assessment as something of a luxury. Moreover, Furtak (2005) indicated that teachers may become satisfied too quickly with their assessment conversations and may also become frustrated by the time and effort necessary to conduct an effective formative assessment.

Furthermore, Alotabi (2014) states that a substantial number of students in a classroom imposes limitations on the classroom activities. Alotabi adds that educators in Saudi Arabia have discovered that oversized classes may affect the application of the new form of assessment in primary schools. High numbers of students in classrooms prevent the teachers from following student progress and providing suitable feedback (ibid). To conclude, this study indicates that training programmes in Saudi Arabia should provide science teachers with the skills and knowledge necessary to implement formative assessment in classrooms because many teachers understand assessment only in terms of providing pupils with marks, indicating their mistakes or moving from one level to the next level. The implementation of this approach represents a substantial task and novel direction in the Saudi and Middle East context, particularly within the existing teaching traditions.

REFERENCES

- Alao, K. A., Kobiowu, S. V., & Adebawale, O. F. (2010). *Fundamentals of educational and counselling psychology*. United Kingdom: Strategies Insight Publishing.
- Alaudan, R. (2015). *Saudi student teachers' perceptions of formative assessment*. Unpublished thesis (PhD), University of York.
- Al-Abdulkareem, S.A. (2004). *Investigating science teachers' beliefs about science and science teaching: struggles in implementing science education reform in Saudi Arabia*. Unpublished thesis (PhD). West Virginia University, Morgantown, USA.
- Alkharusi, H. (2011). Teachers' classroom assessment skills: influence of gender, subject area, grade level, teaching experience, and in-service assessment training. *Journal of Turkish Science Education*, 8 (2), 39-48.
- Alotabi, K.A. (2014). Student assessment strategies in Saudi Arabia: a case study of pre and post classroom practices. *Literacy Information and Computer Education Journal (LICEJ)*, 3 (1), 1786-1763.
- Almazroa, H., & Alshamrani, S. (2015). Saudi science teacher professional development: Trends, practices, and future directions. In N. Mansour and S. Alshamrani (Eds), *Science education in the Arab Gulf States: Visions, sociocultural contexts and challenges* (pp.3-21). Rotterdam: Sense Publishers.
- Al-sadaawi, A.S. (2007). *An investigation of performance-based assessment in science in Saudi primary schools*. Unpublished thesis (PhD), Victoria University.
- Bekiroglu, F.O. (2007). Performance based assessment: theory and practice. *Journal of Turkish Science Education*. 5(1), 113-131.
- Bennett, R.E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18 (1), 5 –25.
- Berry, R. (2008). *Assessment for learning*. Hong Kong: Hong Kong University.
- Black, P., & Wiliam, D. (1998a). Inside the black box: raising standards through classroom assessment. London: King's College.
- Black, P., & Wiliam, D. (1998b). Assessment and Classroom Learning. *Assessment in Education*, 5, (1), 7-74.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning Putting into Practice*. Maidenhead: open University Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2004). working inside the black box :assessment for learning in the classroom . *Phi Delta Kappan*, 86 (1), 8-21.
- Bloom, B. (1984). The search for methods of group instruction as effective as on-to-one tutoring. *Educational Leadership*, 41 (8), 4-17.
- Buck, G., & Trauth-Nare, A. (2009). Preparing teachers to make the formative assessment process integral to science teaching and learning. *Journal of Science Teacher Education*, 20 (5), 475-494.
- Bulunuz, M., & Bulunuz, N. (2013). Formative assessment in science teaching and demonstration of its effective implementation. *Journal of Turkish Science Education*, 10 (4), 119-135.
- Briggs, D.C., & Ruiz-Primo, M.A. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, 31(4), 13-17.
- Carless, D. (2003, November 24-25). Learning-oriented assessment. *a paper presented at the Evaluation and Assessment Conference*,. Adelaide, South Australia.
- Cimer, A. (2007). Effective teaching in science: a review of literature. *Journal of Turkish Science Education*, 4 (1), 20-41.
- Clarke, S. (1998). *Targeting assessment in the primary classroom*. London: Hodder and Stoughton.

- Clarke, S. (2000). Getting it right-distance marking as accessible and effective feedback in the primary classroom. In S. Askew (Ed.), *Feedback for Learning* (pp.32-45). London: RoutledgeFalmer..
- Cole, M.W. (2010). *Influence of assessment for learning professional development in rural Georgia public schools*. Unpublished thesis (PhD). Liberty University.
- Cohen, P.A., Kulik, J.A., & Kulik, C.L.C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19 (2), 237-248.
- Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, A., & Orphanos, S. (2009). *Professional learning in the learning profession: a status report on teacher development in the United States and abroad*. Oxford, OH: National Staff Development Council.
- Doran, R. L., Lawrenz, F., & Helgeson, S. (1994). Research on Assessment in Science. In D. Gabel (Ed.), *Handbook of Research on Science Teaching and Learning* (pp.388-442). New York: MacMillan.
- Dunn, K., & Mulvenon, S. (2009). A Critical Review of Research on Formative Assessment: The Limited Scientific Evidence of the Impact of Formative Assessment in Education. *Practical Assessment, Research, & Evaluation*, 14(7), 1-11.
- Earl, L. (2006). Assessment – a powerful lever for learning. *Brock Education*, 16 (1), 1-15.
- Elstgeest, J. (2001). The right question at the right time. In W. Harlen, (ed.), *primary science: Taking the Plunge*, (2 nd edn, pp. 25-47) Portsmouth, NH: Heinemann.
- Flick, U. (1998). *An Introduction to qualitative research*, London: SAGE.
- Fox-Turnbull, W. (2006). The Influences of teacher knowledge and authentic formative assessment on student learning in technology education. *International Journal of Technology and Design Education*, 16(1), 53-77.
- Frowe, I. (2005). The formative use of assessment information in planning-the notion of contingent planning. *British Journal of Educational Studies*, 53 (1), 54-65.
- Furtak, E.M. (2005). *Formative assessment in K-8 science education: A conceptual review: Commissioned paper by the National Research Council for Science Learning K-8 consensus study*.
- Gibson, K., & Shaw, C.A. (2010). Assessment of active learning. Retrieved from <http://webs.wichita.edu/depttools/depttoolsmemberfiles/carolynshaw/Gibson%20Shaw%20compendium.pdf>. Accessed: 3th August 2014]
- Gioka, O. (2009). Teacher or examiner? The tensions between formative and summative assessment in the case of science coursework. *Research in Science Education*, 39(4), 411-428.
- Kingston, N., & Nash, B. (2011). Formative assessment: a meta-analysis and a call for research. *Educational Measurement. Issues and Practice*, 30(4), 28-37.
- Harlen, W., & James. M. (1996, April). Creating a positive impact of assessment in learning. *Paper presented American Educational Research Association*, New York.
- Harlen W., & Qualter. (2007). *The teaching of science in primary schools*. (4th Ed) London: David Fulton Publishers.
- Hyde, A., Lohan, M., & McDonnell, O. (2004). *Sociology for health professionals in Ireland*. Dublin: Institute of Public Administration.
- Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy. *Education Policy Analysis Archives*, 13(10), 1-28.
- Lambert, D., & Lines, D. (2000). *Understanding assessment: Purposes, perceptions, practice*. New York: Routledge Falmer.

- Meisels, S., Atkins-Burnett, S., Xue, Y., & Bickel, D. D. (2003). Creating a system of accountability: the impact of instructional assessment on elementary children's achievement scores. *Educational Policy Analysis Archives*, 11(9), 1-18.
- Melmer, R., Burmaster, E., & James, T. K. (2008). Attributes of effective formative assessment. Washington, DC: Council of Chief State School Officers.
- National Foundation for the Improvement of Education, (NFIE). (1996). *Teachers take charge of their learning: transforming professional development for student success* Washington, D.C.: NFIE.
- Rodriguez, M.C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education*, 17(1), 1-24.
- Ruiz-Primo, M. A., & Furtak, E.M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11, (3-4), 205-235.
- Sadler, D. (1989). Formative and the design of instructional systems. *Instructional Science*, 18 (2), 119-144.
- Shepard, L.A. (2006). Classroom assessment. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 624-646). Westport, CT: Praeger.
- Stephens, A. (2011). *Formative Assessment: Understanding Learning in the Middle School Classroom*. Unpublished Thesis (MS Education). John Fisher College.
- Task Group on Assessment and Testing, (TGAT). (1988). *National curriculum: a report Task Group on Assessment and Testing*. London: Department of Education and Science/Welsh Office.
- Trumbull, E., & Gerzon, N. (2013). *Professional development on formative assessment: Insights from research and practice*. San Francisco, CA: WestEd.
- Panizzon, D., & Pegg, J. (2008). Assessment practices: empowering mathematics and science in rural secondary schools to enhance student learning. *International Journal of Science and Mathematics Education*, 6 (2), 417-436.
- Phelan, J., Choi, K., Vendlinski, T., Baker, E. L., & Herman, J.L.. (2009). *WWC Review of the Report "The Effects of POWERSOURCE Intervention on Student Understanding of Basic Mathematical Principles"* Retrieved from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=WWCSSR10051> [Date Accessed: 3th August 2014]
- William, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree.
- William, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles Policy and Practice*, 11(1), 49-65.
- Yin, Y. (2005). *The influence of formative assessments on student motivation, achievement, and conceptual change*. Unpublished thesis (PhD), Stanford University.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., & Young, D.B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, 21(4), 335-35.

Teacher-Student Interactions, Learning Commitment, Learning Environment and Their Relationship with Student Learning Comfort

Che Nidzam Che Ahmad¹ , Saidatul Ainoor Shaharim² & Mohd Faizal Nizam Lee Abdullah³

¹Assoc.Prof. Dr., Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Perak, Malaysia

²Mrs., Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Perak, Malaysia

³Assoc. Prof. Dr., Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Perak, Malaysia

Received: 29.02.2016

Revised: 20.10.2016

Accepted: 01.12.2016

The original language of article is English (v.14, n.1, March 2017, pp.57-72, doi: 10.12973/tused.10190a)

ABSTRACT

The purpose of this study was to identify the level of teacher-student interaction, the suitability of learning environment, learning commitment and learning comfort among Biology students. In addition, the contribution of teacher-students interaction, learning commitment and learning environment towards learning comfort were also explored. This quantitative study employed a survey method whereby four types of questionnaires were used: 1) Teacher-Student Interaction Inventory, 2) Classroom Learning Environment Inventory, 3) Learning Commitment Inventory, and 4) Learning Comfort Inventory. A total of 400 Form Four Biology students from Perak were involved in this study. Data were analysed using descriptive and inferential analysis. The findings showed that the level of teacher-student interaction, the suitability of learning environment, learning commitment and learning comfort were high. There were significant relationship between teacher-student interaction, the suitability of learning environment and learning commitment with learning comfort. In addition, the results of multiple regression analysis revealed that students' learning commitment and learning environment provide significant contribution towards learning comfort.

Keywords: Teacher Student interaction, learning commitment, learning environment, learning comfort.

INTRODUCTION

The peaceful and comfortable school atmosphere are very important to students and teachers to ensure effective teaching and learning. This is because students spend most of their time in school. Teachers need to be more creative and innovative in carrying out teaching and learning approaches or strategies so that students could acquire knowledge effectively. According to Akçayl & Doymuş (2014), the selection of appropriate teaching methods and techniques are vital to ensure students' understanding of issues and concept at the highest level. In the classroom, there are various factors that determine the success of students' learning. Teachers, students and environment are among the determining factors that



contribute to the success of the learning process in the classroom. Thus, it is argued that evaluation of all aspects in the learning environment is crucially important because by doing so will not merely providing information for measuring the learners' performance, but also information on teachers' competencies in planning positive learning outcomes (Che Ahmad, 2010).

Good interaction between teachers and students will create positive relationships in the classroom and contribute to effective learning. In fact, effective teachers can assess changes in students' behaviour and understand the needs of students in the classroom. According to Bucholz and Sheffler (2009), teachers can shape the classroom environment to be comfortable and therefore improve the ability of students to learn. Conducive classroom environment emphasizes cooperation and openness between teachers and students. Therefore, in learning and teaching especially science subjects, students should be allowed to interact with teachers, classmates and surroundings. Interaction with the environment during learning process will stimulate all students' senses and encourage students to learn in a creative and innovative way. This phenomenon will trigger students' thinking and refrain them from simply memorizing fact and knowledge. Furthermore, interesting learning experiences will stimulate students' interest in science. Hence, the students will be excited to do various of active and encouraging learning activities. Active involvement in the learning process such as performing an investigation to solve a problem will provide more positive impact on students. According to Ergin, Kanli, & Ünsal (2008) effective teaching by teachers enable students to establish a relation between previous experience and daily life, to apply the information learned in solving problems, to defend their opinion and to take responsibility during his life. Thus, the teacher is a key factor in determining whether students' learning is a success or a failure. According to Telli et al. (2007), there is a general argument that teacher-student interpersonal behaviour is a crucial element in the teaching learning process in any country. Several other studies also showed that teachers' behaviour such as compliments and punishment have a strong impact on students' learning (Walberg, 1984).

Besides teachers, the classroom environment also influences the learning and teaching process. According to Che Ahmad (2013), the physical aspects (size and shape, space, interior light, colour, thermal conditions, noise levels, furniture and seating arrangements, technological facilities) and psychosocial aspects (interaction between students, students and teachers, students and environment) of classroom will contribute to teaching and learning satisfaction and have a significant impact on students' learning. Furthermore, the learning environment and classroom management can enhance productivity and comfort among teachers and students. Comfortable learning is important as discomfort feeling could affect students' motivation to learn. Therefore, classroom learning environment should meet the needs of teachers, students and learning activities conducted, able to stimulate learning, and encourage active participation and more importantly it could be easily facilitated and monitored by the teacher. Bartlett (2003) noted that the effectiveness of learning will be increased if the learning environment is comfortable. Hence, a conducive learning environment, a good student teacher relationship and high learning motivation will increase the learning comfort and improve student achievement over time. This is a very important aspect to be emphasized through the school.

In Malaysia, one of the important classroom is the biology classroom. Biology classroom consist of good students in terms of achievement and the biology learning is important in increasing the students' interest to pursue their studies in medicine. Therefore, the biology classroom environment including the physical properties, teachers and students are very crucial to explore. Hoping by knowing factors that contribut to comfort learning, we might be able to increase their interest and motivation to learn dan consequently more successful in their learning.

Problem Statement

Malaysia Government targeted to produce 60% science stream students has not yet reached its goal (MOE, 2012). Study by Lee and Kamisah (2013) revealed that students' interest in science were declining. This phenomenon occurred not only in Malaysia but also in the whole world (PISA, 2009). There are various possible factors that contribute to the issue of declining student interest in science. Among them are the teacher factor, learning environment, learning motivation and learning commitment.

Teacher-student interaction is essential to establish good relationship between teacher and student. However, Den Brok, Fisher and Scott (2005) and Dhindsa (2006) argued that teachers dominated the interactions in the classroom. According to Kamaruddin (2007), a total of 60.7% of students said that teachers rarely treat students friendly and the mastery of interpersonal skills and communication practices between teachers and students was unsatisfactory. Besides that, the learning environment is also crucial for students' learning. Research showed that changes in local climate will affect microenvironment, such as plants, people and buildings including a school environment and might causes discomfort during the learning process in the classroom, especially for teachers and students (Puteh et al., 2014). In relation to that, Khadijah and Azimin Samson (2013) found that most classrooms were not comfortable and did not meet the needs of teachers and students. Uncomfortable classroom environment can cause many discipline problems and affect effective learning. According to Sahin, Tantekin-Erden & Root, (2011), there were many studies that associated with classroom management and physical environment in the school however, the focus of these studies only focused on one dimension only, such as class size or the physical environment and academic achievement. Not many studies were carried out considering aspects such as student involvement and learning comfort even though it is argued by Cavanagh (2012) that student involvement in learning was also an important consideration in the assessment of the learning environment and instructional design. Furthermore, there were not many studies done combining several aspects and their impact on students. Thus, there was a need to explore several elements that contribute to the learning such as teacher-student interaction, learning environment and learning commitment.

Conceptual Framework

Classroom climate can affect the students' learning comfort. According to Moos (1979), there were four factors that can contribute to a classroom climate. These factors were the characteristics of teachers, physical characteristics and architecture, the student aggregate and organizational factors. All of these four factors were related to each other in forming classroom climate. Based on the Moos model (1979), a conceptual framework was constructed. Within this conceptual framework (Figure 1.1), the inclusion of teacher-student interaction represents the teachers characteristics, the learning environment represents physical features and architecture and student commitment represents the learning aggregate. Components of organizational factors are removed from the conceptual framework with regard that this study does not involve changes to the existing learning environment.. The comfort level of learning is included as outcomes in describing the learning atmosphere that occurs in the classroom climate.

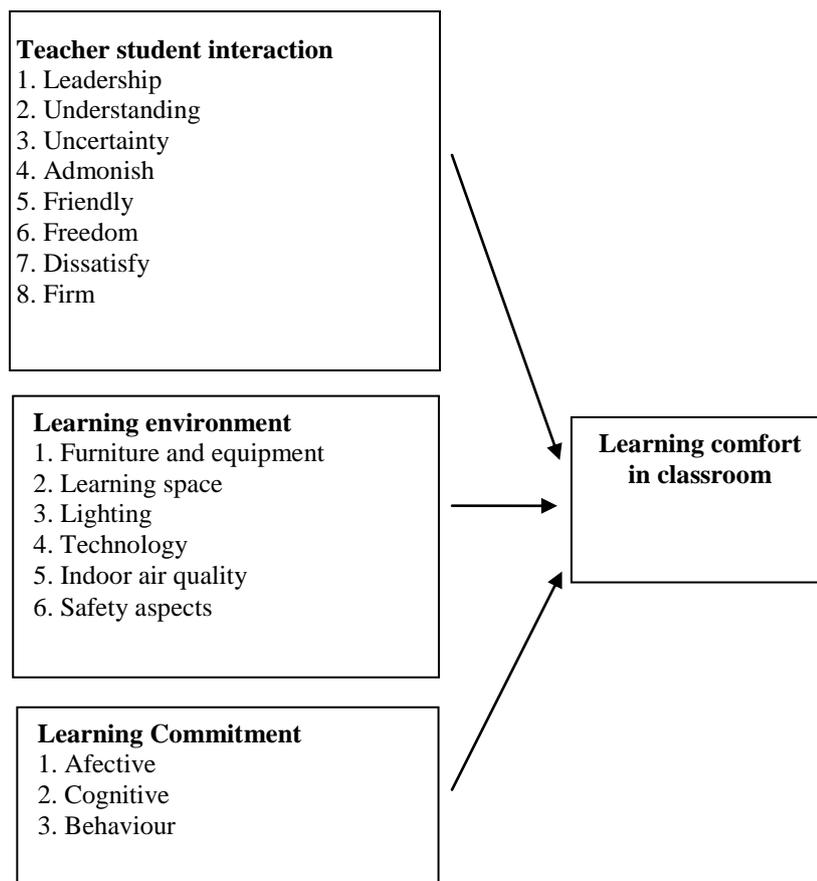


Figure 1 *Conceptual Framework*

Objectives:

- i. Identify the level of teacher-student interaction, the suitability of the learning environment, students learning commitment and classroom learning comfort among Biology students.
- ii. Identify the relationship between teacher-student interactions, the suitability of the learning environment and students leaning commitment with learning comfort among Biology students.
- iii. Identify the contribution of the teacher-student interaction, the suitability of the learning environment and commitment towards learning comfort among Biology students.

METHODS

a) Population and sample

The target population of this study is the form four biology students in Larut Matang and Selama District, Perak, Malaysia. Overall, there were 15 secondary schools located in the urban area and 12 secondary schools in rural areas. The form four students were selected due to the fact that they have been in school for fer years already and are considered mature enough to give realible feedback (Che Ahmad et al, 2010).

For the sample, the researchers used stratified ramdom sampling to choose 10 school in rural area and 10 school in urban area. Twenty biology students randomly selected from each school. Therefore, a total of 400 biology students were involved in this study. The student

are of 165 (41, 2%) were males and 235 (58.8) were females. Approximately 196 (48.92%) students were from rural area and 204 (51.08%) students were from urban school.

b) Instruments

The study used four questionnaires to measure classroom climate that were 1) I Teacher student Interaction Inventory (TSI), 2) Physical Classroom Learning Environment Inventory (PCLEI), 3) Learning Commitment Inventory (LCI) and 4) Students' Learning Comfort Inventory (SLCI). These questionnaires are in the form of a five-point Likert scale with 1 for strongly disagree, 2 for disagree, 3 for less disagree, 4 for agree and 5 for strongly agree.

The first inventory, the Teacher-student Interaction Inventory (TSI), adapted from Questions of Teacher Interaction (QTI) by Wubbels and Levy (1993) was employed to obtain students' view about the teacher-student interaction in the Biology classroom. The TSI consists of eight constructs: leadership, understanding, uncertainty, admonish, friendly, freedom, dissatisfaction and firm. Each scale consists of six items and the entire instrument has 48 items.

The second inventory used was the Physical Classroom Learning Environment Inventory (PCLEI), adapted from the Physical Sciences Laboratory Environment Inventory (PSLEI) developed by Che Ahmad (2011) to identify the suitability of the physical aspects science laboratory learning environment. Amendment was made whereby the word 'science laboratory' in PSLEI was changed to 'classrooms' in PCLEI. PCLEI consists of six constructs: furniture and equipment, learning space, lighting, technology, indoor air quality and safety aspects. Each scale consists of five to eight items and the entire inventory has 28 items.

The third inventory, the students' learning commitment inventory (SLCI), adapted from Rohani et al. (2010) was used to identify the learning commitment among students in the Biology classroom. SCLI consists of three constructs namely cognitive, affective and behavioural. Each construct consists of four to eight items and the entire instrument has 16 items. The fourth inventory, the learning comfort inventory (LCI) was adapted from the scale of Enjoyment of Science Lessons in the Test Of Science - Related Attitudes (TOSRA) by (Fraser & Walberg, 1981) to gather information on students' view concerning student's comfort in the Biology classroom. LCI consists of nine item. The TSI, SLCI and LCI are originally written in English, and for this research, all these three inventories were translated into Bahasa Malaysia using the back to back translation procedure as recommended by Brislin (1976) and Newmark (1988) to make sure that the Bahasa Malaysia questionnaires conveyed the same meaning as the English questionnaires.

The validity of these inventories, TSI, CLEI, SLCI and LCI were determined by three experts in their respective fields. The experts' percentage agreement for all items in the inventories used (TSI, PCLEI, SLCI and LCI) were more than 80% as required to be valid (Hamzah et al., 2013). The inventories were amended and improved according to the comments made by the experts. Then, a pilot study was conducted to check the reliability of the questionnaires. The reliability (Cronbach alpha coefficient) for the TSI, PCLEI, SLCI and LCI ranged from 0.65 to 0.92. This range was regarded as acceptable to good (George and Mallery 2001; Hair et al. 2006, 2010), because the closer alpha is to 1, the greater the internal consistency of the items. Hence, the reliability of the TSI was confirmed for this sample.

FINDINGS

a) Teachers students interaction

The mean value of teacher-student interaction from the perspective of the students are shown in Table 1. Overall, the findings indicate that the understanding construct has the highest level of practice (mean = 4.37, SD = 0.62), followed by a friendly construct (mean = 4.29, SD = 0.62), admonish construct (Mean = 4.28, SD = 0.84), uncertainty construct (mean = 4.27, SD = 0.77), dissatisfaction constructs (mean = 4.23, SD = 0.87), leadership construct (mean = 4.22, SD = 0.67), firm construct (mean = 3.99, SD = 0.66) and freedom construct (mean = 3.32, SD = 0.89)

Table 1. Mean value and standard deviation for teacher student interaction

No.	Construct	Mean	SD	Interpretation	Level of practice
1.	Leadership	4.22	0.67	Positive	High
2.	Understanding	4.37	0.62	Positive	High
3.	Uncertainty	4.27	0.77	Positive	High
4.	Admonish	4.28	0.84	Positive	High
5.	Friendly	4.29	0.62	Positive	High
6.	Freedom	3.32	0.89	Negative	Moderate
7.	Dissatisfaction	4.23	0.87	Positive	High
8.	Firm	3.99	0.66	Positive	High

Findings also show that the constructs of leadership, understanding, friendly, uncertainty, admonish, firm and dissatisfaction are positive. This shows that students perceive their teachers of having a high level of leadership whereby they are able to draw students' attention, care for students, friendly assists the students and exhibit firmness to ensure students conform to the rules during the learning process. Students also perceive that their teachers look confident, patient in advising student and also exhibit satisfaction during learning conducted in the classroom. On the other hand, students perceive freedom construct practices are at moderate level meaning that the students assume teacher of providing them less opportunities to take over the responsibility of their own learning.

b) The suitability of learning environment

The mean value of the suitability of the learning environment from the perspective of the students are shown in Table 2. Overall, the findings indicate that the learning space construct have the highest level of suitability (mean = 4.28, SD = 0.57), followed by lighting constructs (mean = 4.18, SD = 0.62), furniture and equipment construct (mean = 3.86, SD = 0.73), safety aspects (mean = 3.68, SD = 0.76), technology construct (mean = 3.12, SD = 1.06) and indoor air quality construct (mean = 2.23, SD = 0.80)

Table 2. Mean value and standard deviation for the suitability of learning environment

No.	Construct	Mean	SD	Interpretation
1.	Furniture and equipment	3.86	0.73	High
2.	Learning space	4.28	0.57	High
3.	Lighting	4.18	0.62	High
4.	Technology	3.12	1.06	Moderate
5.	Indoor air quality	3.54	0.80	Moderate
6.	Safety aspects	3.68	0.76	High

The findings show that the learning space, lighting, furniture and equipment, and safety aspect have high level of suitability from students' perspective meaning that students thought

learning space, lighting, furniture and equipment, and safety aspects provided in the classroom is suitable and can be adapted with various learning activities in the classroom.

c) The Biology students' learning commitment

The mean value and the level of students' learning commitment are shown in Table 3. As a whole, the findings indicate that students have a high level of commitment to learning in all three sub-constructs which are behaviour (Mean = 4.14, SD = 0.67), followed by cognitive constructs (Mean = 4.12, SD = 0.60) and affective construct (Mean = 4.04, SD = 0.67).

Table 3. Mean value and standard deviation for students' learning commitment

No.	Construct	Mean	SD	Interpretation	Level of practice
1.	Cognitive	4.12	0.60	Positive	High
2.	Affective	4.04	0.78	Positive	High
3.	Behaviour	4.14	0.67	Positive	High

Students rate their learning commitment as positive in the aspect of cognitive, affective and behaviour. This shows that students' learning commitment is higher than their thoughts, emotions and physical actions. The findings of this study suggest that students are very committed to the learning in the classroom

d) Comfort learning

The mean value and comfort level of learning from students' perspective are; mean, 3.99 and SD, 0.73 as shown in Table 4.

Table 4. Mean value and standard deviation for learning comfort

No.	Construct	Mean	SD	Interpretation
1.	Learning comfort	3.99	0.73	Positive

The learning comfort construct has a positive level from the view of students. This shows that students are comfortable with the learning in the classroom. The findings indicate that the process of learning and teaching biology in the classroom are at a pleasant situation. Students feel very satisfied with the learning environment in the classroom.

e) Relationship between teacher student interaction and comfort learning

Table 5 shows the results of Pearson correlation between teacher-student interaction and learning comfort.

Table 5. Pearson correlation between teacher-student interaction and learning comfort.

Variable	1	2	Sig. (2-tailed)
Teacher student interaction (1)		0.568*	0.000
Learning comfort (2)			

*significant at level $p < 0.05$

Based on the Table 5, analysis reveals that the teacher-student interaction correlate significantly with the learning comfort in the classroom ($r = 0.568$; $p < 0.05$). This shows that there is a strong positive relationship between teacher-student interactions with the learning

comfort. Its means that when interaction between teachers-students are good, the learning comfort will exit.

f) Relationship between the suitability of learning environment interaction and comfort learning

Table 6 shows the results of Pearson correlation between the suitability learning environment and learning comfort.

Table 6. *Pearson correlation between the suitability learning environment and learning comfort*

Variable	1	2	Sig. (2-tailed)
Learning environment (1)		0.359**	0.000
Learning comfort (2)			

** Significant at level $p < 0.01$

Based on the Table 6, the analysis shows that the suitability of learning correlate significantly with the learning comfort in the classroom ($r = 0.359$; $p < 0.01$). This shows that there exists a simple positive correlation between learning environment with learning comfort. In addition, it indicates that the suitability of the learning environment have a significant relationship with the comfort learning in Biology classroom meaning that when the learning environment is suitability, in good condition and conducive, students will feel comfortable to learn in the classroom.

g) Relationship between learning commitment and learning comfort

Table 7 shows the results of Pearson correlation between learning commitment and learning comfort.

Table 7. *Pearson correlation between learning commitment and learning comfort.*

Variable	1	2	Sig. (2-tailed)
Learning commitment (1)		0.563**	0.000
Learning comfort (2)			

** Significant at level $p < 0.01$

According to Table 7, analysis revealed that the learning commitment correlate significantly with the learning comfort in the classroom ($r=0.563$; $p < 0.01$). This shows that the learning commitment to learn has a significant relationship with the learning comfort in biology classroom. This means that when students' committed in their learning, then students will be more comfortable with the learning and teaching in the classroom.

h) Contribution of teacher student interaction, suitability of learning environment and Biology learning commitment towards learning comfort

Table 8 and Table 9 shows the results of multiple regression analysis for the three predictor variables on Biology learning comfort. Based on Table 8, the analysis shows that there are two predictor variables that have a significant correlation with the Biology learning comfort $F(2,309) = 261.90$ with significant value = 0.000 ($p < 0.05$). Therefore, there are significant contributions by the predictor variables (learning commitment and learning environments) with biology learning comfort.

Table 8. Multiple regression analysis

Model	Variation	Total power of two	df	Mean Squares	F	Significant
2	Regression	82.609	2	41.305	261.901	0.000 ^c
	Error	48.733	309	0.158		
	Total	131.342	311			

* Significant at level $p < 0.05$

Table 9. Multiple regression analysis (stepwise)

Predictor	B	Standard error	Beta (β)	T value	Significant level	R ²	Contribution (%)
Constant	0.140	0.182		0.772	0.441		
Learning commitment	0.794	0.050	0.700	15.932	0.000	0.617	61.7
Learning environment	0.173	0.055	0.138	3.147	0.002	0.629	1.2

Dependent: comfort

Double R 0.793

R Square (R²) 0.629

R Square adjusted 0.627

Standard error 0.39713

Contribution 62.9%

Primary and utmost predictor for Biology learning comfort is learning commitment factor which contribute 61.7 percent. This situation shows that for every unit increase in score of learning commitment, Biology learning comfort score will increase by 0.7 units. Second predictor is the learning environment factors which contributed 1.2 percent to the Biology learning comfort. This means that when the learning environment increased per unit, the score of Biology learning comfort will increased 0.138 units.

Results of the analysis show that the two predictor variables, namely learning commitment and learning environment are a predictor for the Biology learning comfort. Value R² = 0.629 indicates the overall contribution of two predictor variables to Biology learning comfort were 62.9 percent. The contribution of learning commitment is 61.7 percent whereas the contribution of learning environment is 1.2 percent. Overall, these two factors are significant to the biology learning comfort. Thus, from the results of multiple regression, the equation for regression is as follows:

Multiple regression model

$$Y = 0.140 + 0.794X_1 + 0.173X_2$$

$$Y = a + b_1X_1 + b_2X_2 + e$$

Y = Biology learning comfort

X₁ = learning commitment

X₂ = learning environment

b₁ and b₂ are the coefficients for the related independent variables related

Constant 0.140

DISCUSSION and CONCLUSION

a) Teachers students interaction

Overall, students perceived that the teacher-student interaction practices was good and effective. This was important because according to Myint and Atputhasamy (2005) the quality of teacher leadership behaviour was an indicator of the quality teacher-student interaction in the classroom. In fact, the teacher-student interaction was an important element to the students in the context of teaching, learning and assessment (Douglas et al., 2015). Fraser et al. (2010) described teachers who want to improve students' academic achievement should show strong leadership behaviour and understand, and reduce the uncertainty behaviour in the classroom. Ab. Samad & Jamaluddin (2005) also explained that the elements of leadership in the classroom as practiced by the teacher can determine the attainment of a classroom and also effect the students. According Wubbels and Levy (1991), teachers' behaviour in a classroom setting is important because it can influence students' motivation and achievement. Students' perceived moderate practices of freedom due to teachers control and monitor learning in the classroom. Monitoring activities should be done to prevent students from becoming too independent and unmanageable. According to Cruickshank et al. (2009) when the teacher gives freedom to the students to learn by themselves and leave the classroom without proper control, the students will be less motivated to learn and are more likely to exhibit negative behaviors such as not completing the task. Therefore, as a teacher, it is acceptable to be more assertive and monitor the learning so that students are serious and devote more attention to learning.

b) The suitability of learning environment

The findings indicated that the learning space, lighting, furniture and equipment, and safety aspect have high level of suitability from students' perspective. In other words students perceived learning space, lighting, furniture and equipment, and safety aspects provided in the classroom is suitable and can be adapted with various learning activities in the classroom. The suitability of this learning space may be due to the wide classroom area that was appropriate to carry out various learning activities. Thus, allowed the students in the classroom to move around freely and perform the learning activities either individually or in groups. Besides that, lighting in classrooms was found to be good and appropriate due to combination of natural and artificial lighting. According to Barnitt (2003), combination of natural light and lamp provide quality lighting. Furthermore, the brightness of the classroom can also be controlled with the use of a separate switch, blinds or curtains as required by learning activities conducted. Furniture and equipment and the safety element were also perceived suitable by students. The suitability may be attributed by the improvement of infrastructure facilities and equipment in the classroom studied. This was in line with the government's efforts to upgrade the facilities involved in all schools to support the educational needs of students from time to time (PIPP 2013-2025). However, this contrasted with the findings from Parveen Khan (2012) and Che Ahmad who found that the suitability was at the moderate level and must be addressed in order to provide a positive impact on students' learning and teaching.

Furthermore, students also perceived technology construct at moderate level. It implied that there was insufficient number of computer or no computer in the classroom. Study by Che Ahmad (2011) also found that there were no computers for students' use in the learning space. The classroom should be equipped with technology such as computers and accessible to internet in order to facilitate the search for information during the teaching and learning process. Therefore, technology equipment in classroom needs to be improved from time to time to meet the needs of students. In addition, Rawlins and Kehrwald (2014) argued that ICT

is an effective pedagogical tool and potentially in assisting teachers learning and teaching in classroom. With the flexibility and help of the technologies, teachers can design learning environments in which students can manage and construct their own representations of knowledge in their minds (Koç, 2005). In addition, the learning environment should create a good atmosphere to support the learning and teaching process. According to Zakaria et al. (2012) learning environment can impact students' behaviour and self-esteem. The physical characteristics of learning environment could also influence teacher-student communication and have an impact on cognitive and affective domains (Che Ahmad et al., 2014). Thus, in order to create a conducive learning environment, student-centred and fulfil the needs of teacher and student, the environment must be assessed and improved to enhance the effectiveness of learning. The characteristics of a classroom will impact overall satisfaction (Hill & Kathryn 2010). Therefore, students will be more enjoy and satisfy in a well-designed classroom. In fact, the quality of education not only be assessed by the subjects taught and the level of students' achievement but also the evaluation of the classroom learning environment.

c) The Biology students' learning commitment

Findings also revealed that students rated their learning commitment in the aspect of cognitive, affective and behaviour as positive. This showed that students' learning commitment is higher than their thoughts, emotions and physical actions. The findings of this study suggested that students were very committed to learn in the classroom. These students were very interest in learning and motivated in completing the tasks assigned by the teacher. The findings were in line with Wonglorsaichon et al. (2014) which stated that students who have learning commitment in terms of behavioural, cognitive and affective will produce more quality learning. It was also supported by Covell et al. (2009), which explained that the commitment of student learning will benefit various parties involved, directly or indirectly. According to Mohd Nihra et al. (2012), students who were less commitment likely to abandon a task, less preparing for exams and all sorts of things that can interfere with the ongoing learning and teaching session in the classroom. It seemed that students who fail to adapt to the learning will likely face many problems in terms of academic, social, emotional, and commitment to learning. Students involved and committed to learning when they felt appreciated, understood and engaged in learning that occurs.

d) Comfort learning

Students viewed the learning comfort construct as positive. This showed that students were comfortable with the learning in the classroom. The findings indicated that the process of learning and teaching biology in the classroom are at a pleasant situation. Students felt very satisfied with the learning environment in the classroom. However, Marzita et al. (2014) found that the classroom comfort was at moderate level. This may be due to the study's location which was carried out in urban areas with high population density. The high population density in urban areas may contribute to the high number of students in a class. Too many students will lead to congestion in classroom and thus reduce the comfort of learning. Students allocate most of their time in the classroom. Therefore, comfort is the key in encouraging learning effectiveness. Comfortable learning environment which meet the needs of students will encourage active participation and enhance the understanding of the concept. Learning effectiveness can be strengthened if the physical and psychological comfort is taken into account. The comfort environment are also related with learning productivity and it's also depend on the building's design and pararell with student's activities (Vale'ria Azzi

da GRAC Collet et al., 2007). Thus, according to Weilin et al. (2013) learning comfort could increase students' motivation to learn in the classroom

e) Relationship between teacher student interaction and comfort learning

Analysis showed that there was a strong positive relationship between teacher-student interactions with the learning comfort meaning that when interaction between teachers-students are good, the learning comfort will exist. Teacher-student interaction was the most important element to a successful relationship during school (Robert & Bridget, 2009). Furthermore, two-way communication between students and teachers play an important role in student's academic achievement (Solmaz et al., 2013). The interaction that occurred between teachers and students will increase the learning comfort because students felt that they were heard, understood and appreciated.

f) Relationship between the suitability of learning environment interaction and comfort learning

The learning environment had a significant relationship with the comfort learning in Biology classroom meaning that when the learning environment is suitable, in good condition and conducive, students will feel comfortable to learn in the classroom. This is because the suitability of a learning environment will influence students' acceptance and promote effective learning process. This result is in line with Basey et al. (2008) statement that the environment can positively influence students' attitudes. The availability equipment, rules of safety, management and learning environment can also increase students' productivity and promising comfort in the classroom. Khadijah and Azimin Samson (2013) found that many classrooms did not provide a comfortable learning environment and meet the needs of teachers and students. This situation might reduce students' concentration and commitment in classroom learning sessions. Teachers need to shape the learning environment with regard to the design of the physical environment in terms of learning space equipped with the technology facilities to ensure students' comfort and increase the effectiveness of teaching and learning (Lei, 2010). According to Veal and Jackson (2005), the design of learning environments affects the level of students' interaction and involvement. This is because when the learning environment meet the needs of teachers and learning activities, various strategies and approaches could be used in learning and teaching. In fact, the physical environment of the classroom is an effective catalyst for the learning and therefore producing quality human capital.

g) Relationship between learning commitment and learning comfort

The learning commitment also had a significant relationship with the learning comfort in biology classroom. This means that when students were committed in their learning, then they will be more comfortable with the learning and teaching in the classroom. The findings reflected that students having passion for learning will tend to feel comfortable in learning in order to success in their life. This is supported by Dallimore et al. (2008), who stated that there is a relationship between learning comforts with a students' commitment. Students who are comfortable in their learning will be actively involved in the learning process. This involvement extremely beneficial in improving their discussions, negotiation of meaning and ultimate motivation in wanting to study Biology (Solas & Wilson, 2015). The involvement of students will increase their mastery in content knowledge and further enhance student

learning commitment. In fact, according to Ghorbani et al. (2013) student learning commitment will inculcate discipline and a greater sense of responsibility in them.

h) Contribution of teacher student interaction, suitability of learning environment and Biology learning commitment towards learning comfort

Finally, analysis revealed that learning commitment and learning environments were the two predictor variables that contributed significantly to students' learning comfort in Biology classroom. The finding is supported by Serafimova and Zdravkovska (2013), who argued that in order to increase the motivation, involvement and students' learning commitment, the students' confidence, mutual respect, active communication should be strengthened over time. This is also consistent with Ghafouri (2014) findings whereby students built their own learning experience involving emotion and cognition. In summary, the analysis found learning commitment and learning environments were predictor variables that contribute significantly to the variations of Biology learning comfort. Thus, in order to improve the effectiveness of the learning and teaching, teachers need to use the existing environment wisely in order to guide and educate students and enhance their knowledge.

CONCLUSION

The findings showed that the level of teacher-student interaction, the suitability of learning environment, the student learning commitment and the biology learning comfort were positive. There were also a relationship between student teachers interaction, the suitability of the learning environment and learning commitment with the learning comfort. Further analysis also showed that the two main factors that contributed to the learning comfort were learning commitment and the learning environment. Therefore, teachers should employed multi methods and strategies in teaching biology in a more attractive way in the classroom. In addition, teachers could have group activities to improve communications and interactions among students and also promote active participation during lessons. Student-centred teaching should be emphasized so that students are more confident and motivated to learn in the classroom.

REFERENCES

- Ab. Samad, R.S & Jamaludin, S. (2005). The role of teachers' leadership in influencing the classroom climate. *Makalah pendidikan*, 117-129.
- Akçay, N. O. & Doymuş, K. (2014). The Effect of Different Methods of Cooperative Learning Model on Academic Achievement in Physics, *Journal of Turkish Science Education*, 11(4), 17-30.
- Barnitt, H. 2003. Lighting for the future. *Building Services Journal: The magazine for the CIBSE* 25 (1): 38-39.
- Bartlett, B.A. (2003). Blogging to Learn. *Knowledge Tree E-Journal*. Retrieved from: http://knowledgetree.flexiblelearning.net.au/edition04/pdf/Blogging_olearn.pdf.
- Basey, J., Sackett, L. & Robinson, N. (2008). Optimal science lab design: impacts of various components of lab design on students toward lab. *International Journal for the Scholarship of Teaching and Learning* 2, 1-15

- Bucholz, J.L. & Sheffler, J. L. (2009). Creating a warm and inclusive classroom environment: Planning for all children to feel welcome. *Electronic Journal for Inclusive Education*, 2(4), 1-14
- Brislin, R.W. 1976. Comparative research methodology: Cross-cultural studies. *International Journal of Psychology 11*: 215-229.
- Cavanagh, R. (2012). Associations between the classroom learning environment and student engagement in learning 1: A rash Model Approach. In Knight, J, (Ed), The Joint Australian Association for Research in Education and Asia Pasific Educational Research Association Conference (AARE-APER), Dec 2-6 Sydney, Australia: Australian Association for Research in Education (AARE).
- Che Ahmad, C. N. (2011). Physical and psychosocial aspects in science laboratory learning environment. Doctoral Thesis. Universiti Kebangsaan Malaysia.
- Che Ahmad, C.N, Osman, K & Halim, L. (2010). Predictive Relationship between Science Laboratory Learning Environment and Level of Students Satisfaction. *Jurnal Pendidikan Malaysia*, 35(2) 19-30.
- Che Ahmad, C.N, Osman, K & Halim, L. (2013). Physical and psychosocial aspects of the learning environment in the science laboratory and their relationship to teacher satisfaction. *Learning Environ Research*, 16,367–385
- Che Ahmad, C.N, Osman, K & Halim, L. (2014).The Establishment of Physical Aspects of Science Laboratory Environment Inventory (PSLEI). *Journal of Turkish Science Education*, 11(2), 123-136
- Covell, K., McNeil, J.K. & Howe, R.B. (2009). Reducing teacher burnout by increasing student engagement: A children's rights approach. *Journal of School Psychology International*, 30, 282–290.
- Cruikshank D.R., Jenkins D.B., & Metcalf K.K. (2009). *The Act of Teaching* (5th ed.). Boston: McGraw Hill.
- Dallimore, E. J. and Platt, M and Hertenstein, J. H. (2008). Class Participation in Accounting Courses: Relationship Among Preparation, Frequency, Comfort, and Learning. AAA 2009 Management Accounting Section (MAS) Meeting Paper. Available at SSRN: <http://ssrn.com/abstract=1183882> or <http://dx.doi.org/10.2139/ssrn.1183882>
- Den Brok, P., Fisher, D., & Scott, R. (2005). The importance of teacher interpersonal behaviour for student attitudes in Brunei primary Science classes. *International Journal of Science Education*, 27(7), 765–779.
- Dhindsa, H. S. (2006). Teacher communication behaviour and enjoyment of science lessons. In M. B. Klein (Ed.), *New Teaching and Teacher Issues* (pp. 115-140). New York, NY: Nova
- Douglas, J. A., Douglas, A., McClelland, R. J., & Davies, J (2015) Understanding student satisfaction and dissatisfaction: an interpretive study in the UK higher education context. *Studies in Higher Education*, 40(2), 329-349.
- Ergin, İ, Kanli, U & Ünsal, Y. (2008). An Example for the Effect of 5E Model on the Academic Success and Attitude Levels of Students': Inclined Projectile Motion. *Journal of Turkish Science Education*, 5 (3), 47-59.
- Fraser & Walberg, 1981 Fraser. B. J. & Walberg, H.J. (1981). Psychological learning environment in science classrooms: A review of research. *Studies in Science Education* 8: 67-92.
- Fraser, B. J., Aldridge, J. M., & Soerjaningsih, W. (2010). Instructor-student interpersonal interaction and student outcomes at the university level in Indonesia. *The Open Education Journal*, 3, 21–33.
- George, D., & Mallery, P. (2001). *SPSS for windows step by step: A simple guide and reference 10.0 update* (3rd ed.). Toronto: Allyn and Bacon

- Ghafouri, F. (2014). Close encounters with nature in an urban kindergarten: A study of learners' inquiry and experience. *Education 3-13: International Journal of Primary, Elementary and Early Years Education*, 42(1), 54-76.
- Ghorbani Somayeh, Jaafari SayyedMostafa & Arbabisarjou Azizollah. (2013). Investigating the effect of positive discipline on the learning process and its achieving strategies with focusing on the students' abilities. *International Journal of Academic Research in Business and Social Sciences*, 3(5), 2222-6990.
- Hair, J.F. Jr., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006). *Multivariate Data Analysis* (6th Edition). Upper Saddle River, NJ: Prentice Hall.
- Hair, J.F. Jr., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2010). *Multivariate Data Analysis* (7th Edition). Upper Saddle River, NJ: Prentice Hall.
- Hill, M. C., & Epps, K. K. (2010). The impact of physical classroom environment on student satisfaction and student evaluation of teaching in the university environment. *Academy of Educational Leadership Journal*, 14(4), 65-79.
- Hamzah, M.S.G., Paim, L., haron, S. A., & Abdullah, M. F.N. L. (2013). *Buku Panduan penggunaan instrumen: anda dan kepenggunaan*. Emeritus Publications, Tanjung Malim, Perak.
- Kamaruddin, K. (2007). Secondary school teachers' interpersonal skills. *Jurnal Pedagogi*, 1(2), 47-62
- Khadijah Kamaruzzaman & Azimin Samsul Mohd Tazilan. (2013). Thermal comfort assessment of a classroom in tropical climate conditions. *Recent Advances in Energy, Environment & Development*, 30(1), 87-91.
- Khan, P., & Iqbal, M. (2012). Role of physical facilities in teaching learning process. *Interdisciplinary Journal of Contemporary Research in Business*, 4 (3), 210- 218
- Koç, M. (2005). Implications of Learning Theories for Effective Technology Integration and Pre-service Teacher Training: A Critical Literature Review. *Journal of Turkis Science Education*, 2 (1), 2-18.
- Krejcie, R.V. & Morgan, D.W. (1970). Determining sampel size for research activities. *Educational and Psychological Measurement*, 30(3), 607-610.
- Lee, T. T & Kamisah Osman. (2013). Impact of interactive multimedia module with pedagogical agents on students's understanding and motivation in the learning of Electrochemistry. *International Journal of Science and Mathematics Education*, 12 (2), 395-421
- Lei, S. A. (2010). Classroom physical design influencing student learning and evaluations of college instructors: A Review of Literature. *Education*, 131(1), 128-134.
- Marzita Puteh, Mazlini Adnan, Mohd Hairiy Ibrahim, Noraini Mohamed Noh & Che Nidzam Che Ahmad. (2014). An analysis of comfortable teaching and learning environment: Community response to climate change in school. *Procedia - Social and Behavioral Sciences*, 116, 285 – 290.
- Mohd Nihra Haruzuan Mohd Said & Chandrathevy Kuppusamy. (2012). Hubungan sikap (komitmen, emosi, berwawasan) dengan pencapaian akademik pelajar PKPG (SPT, SPN, SPH) tahun akhir di Fakulti Pendidikan, UTM Skudai. *Journal of Educational*, 6, 74–91.
- Moos, R.H. (1979). *Evaluating Educational Environments: Procedures, Measures, Findings and Policy Implications*. San Francisco: Jossey-Bass.
- Myint, S. K. & Lourdusamy Atputhasamy. (2005). Self-perceived and students' perceptions of teacher interaction in the classrooms. A paper presented at the Conference on Redesigning Pedagogy; Research, Policy, Practice, Singapore, 30 May - 1 June.
- Newmark, P. 1988. *A Textbook of Translation*. NJ: Prentice Hall
- Kementerian Pendidikan Malaysia (2013). Plan Induk pembangunan Pendidikan 2013-2025.

- PISA. (2009). *Assessment Framework: Key competencies in Reading, Mathematics and Science*. OECD Publishing.
- Rawlins, P. & Kehrwald, B. (2014). Integrating educational technologies into teacher education: A case study. *Innovations in Education and Teaching International*, 51(2), 207-217.
- Robert, C. P & Bridget K. H. (2009). New Directions for Youth Development. *Spring 2009*, 121, 33-46.
- Rohani Arbaa, Hazri Jamil & Nordin Abd Razak, (2010), Hubungan guru-pelajar dan kaitannya dengan komitmen belajar pelajar: Adakah guru berkualiti menghasilkan perbezaan pembelajaran antara jantina pelajar? *Jurnal Pendidikan Malaysia*, 35(2), 61-69.
- Sahin, I.T., Tantekin-Erden, F., & Akar,H. (2011). The influence of the physical environment on early childhood education classroom management. *Egitim Arastirmalari-Eurasian Journal of Education Research*, 44,185-202.
- Solas, E. C. & Wilson, K. (2015). Lessons Learned and Strategies Used While Teaching Core-Curriculum Science Courses to English Language Learners at A Middle Eastern University. *Journal of Turkish Science Education*, 12(2), 81-94
- Solmaz Abdoirahmi Javid, Adei Zahed-Babelan & Yusef Namvar. (2013). A study on the state of teacher-student verbal interactions during teaching process and its relationship with academic achievement of middle school students in Ardabil. *International Research Journal of Applied and Basic Sciences*, 4(7), 1909-1913.
- Suzana, D, Z & Serafimova, Mimoza, S. (2013). Communication Beyond the lecture room as a challenge of education in the modern ERA. *Journal of Sustainable Development*, 4 (7), 5-19
- Telli, S., den Brok, P & Æ Jal Cakiroglu, J. (2007). Students' perceptions of science teachers' interpersonal behaviour in secondary schools: Development of a Turkish version of the Questionnaire on Teacher Interaction, *Learning Environ Research*, 10,115–129
- da Grac, V. A.C., Kowaltowski, D.C.C.K., & Petreche. J.R.D. (2007). An evaluation method for school building design at the preliminary phase with optimisation of aspects of environmental comfort for the school system of The State Sao Paulo in Brazil. *Building and Environment*, 42 (2007), 984–999.
- Veal, W.R & Jackson,Z. (2005). Developing a primary Science methods classroom. *International Journal of Science and Mathematics Educations*, 4, 195-213.
- Walberg, H.J. (1984). Improving the productivity of America's schools. *Educational leadership (Alexandria, VA)*, 41(8), 19 –27.
- Weilin, C., Guoguang, C., Jung, H.P., Qin, O., Yingxin, Z., (2013). Influence of indoor air temperature on human thermal comfort, motivation and performance. *Building and Environment*, 68(2013), 114-122.
- Wonglorsaichon, B., Wongwanich, S., & Wiratchai, N. (2014). The influence of students' school engagement on learning achievement: A Structural Equation Modeling Analysis. *Procedia-Social and Behavioral Sciences*, 116), 1748-1755.
- Wubbels, T. & Levy, J. (1991). A comparison of interpersonal behavior of Dutch and American teachers. *International Journal of Intercultural Relations*, 15, 1-18.
- Zakaria Stapa, Ahmad Munawar Ismail & Noranizah Yusuf. (2012). Social environmental factors and their relation to identity formation. *Jurnal Hadhari Special Edition*, 155–172.

Analysis of Scientific Language of Household Cleaning Products’ Labelling and Its Educational Implications

Angel EZQUERRA¹ , Belen FERNANDEZ-SANCHEZ², Marina MAGAÑA³, Beatriz MINGO⁴

¹Prof. Ph.D., Faculty of Education, Complutense University of Madrid, Spain

²Ph.D., Faculty of Education, Complutense University of Madrid, Spain

³Prof. Ph.D., Faculty of Education, Complutense University of Madrid, Spain

⁴Ph.D. Student, Faculty of Chemistry, Complutense University of Madrid, Spain

Received: 29.02.2016

Revised: 20.10.2016

Accepted: 01.12.2017

The original language of article is English (v.14, n.1, March 2017, pp.73-88, doi: 10.12973/tused.10191a)

ABSTRACT

The necessity of an adequate scientific literacy among the population in general and students in particular is a fact. Two of the main obstacles in science education that hinder it are the lack of connection between the contents worked in class and the students’ day-to-day, and the misunderstanding of scientific language. In the present study, a categorization and analysis of the scientific content of labelling and packaging of household cleaning products is presented. For that, we collected information from more than 500 items from 10 different countries and analysed it through a methodology based on a content analysis of a written text. According to the criteria used –the comparison of each of these contents with the different contexts and knowledge set by the OECD in the 2015 PISA Science Framework- we classified and analysed the information collected according its type of information, obtaining different categories: information referred to physical units, composition of the product, procedures, security instructions, environmental specifications and advertising information. The results and discussion describe the mentioned categories in which this information is classified, and regarded to the type, function, and form of language used in each one. Finally, some important social implication and consequences concerning to the language used in household cleaning products are highlighted in the conclusions. Complementarily, some general recommendations of how to work these contents in class are sketched out.

Keywords: scientific literacy, scientific language, labelling, household cleaning products

INTRODUCTION

Since science and technology are completely immersed in our lives, it could be unimaginable to conceive our day-to-day without all the scientific and technological developments. From the satellites that take pictures of the Earth, to the smartphones that allow us to check it. These advances have drastically changed different areas of our society, which include communication, transportation, medicine and leisure, among others. Their influence within the household environment is also evident, where its implementation is mostly related to the simplification of the domestic chores. For



instance, in the 19th century, laundry was done by the riverside using a wooden washboard and a bar of soap, while now it is easily done at home with a washing machine and detergent.

Given the great amount of scientific contents present in our daily lives, population has to have a minimum understanding of the basic scientific concepts and processes in order to act coherently and successfully face the situations where science and technology are involved. This is known as scientific literacy (Miller, 1983; Kolstø, 2006; Lewis & Leach, 2006). Achieving the scientific literacy among the citizenship is absolutely necessary due to a wide range of reasons: economic -scientific and technological development of a country or region-; political and social -democratic understanding and participation in science and technology issues-; cultural -understanding the scientific knowledge as an essential part of the culture-; and functional -necessity to understand and cope with today's world- (DeBoer, 2000; Hodson, 2003; Bybee, 2008). Accordingly, the scientific literacy term can be analysed from two perspectives: the macro, related to a global benefit for the nation or society, and the micro, which concerns the improvement of the individual's quality life (Laugksch, 2000); in other words, under both the social and the personal point of view (Harlen, 2001).

The growing concern of society in achieving the scientific literacy has resulted in the publication of a high number of declarations and congresses considering this issue (e.g. OECD, 2007; UNESCO 2008). Specifically, the European Parliament approved the European Framework for Key Competences for Lifelong Learning that established eight key competences for lifelong learning, which includes the scientific competence (EC, 2006). Furthermore, in formal education, most European schools curricula follow this educational perspective with the aim of succeeding in the acquisition of a proper scientific literacy from the early educational stages (COSCE, 2011).

However, despite the numerous efforts made by the educative community, the current situation in science education does not meet the proposed objectives. Concretely, there are a high percentage of students who are quite far from succeeding in the acquisition of the adequate level of scientific literacy (Osborne & Dillon, 2008; Martin, Mullis, Foy & Stanco, 2012). There are several reasons that could explain such results; however, two of the main difficulties that have been identified in science education are the lack of motivation of students (Tuan, Chin & Shieh, 2005) and miscomprehension of scientific language (Mercer, Dawes, Wegerif & Sams, 2004; Aikenhead, 2005; Sjøberg & Schreiner, 2010; Osborne & Collins, 2001).

Among other reasons, students' lack of motivation might be associated to the disconnection between topics studied in class and their everyday life (Duggan & Gott, 2002; Stuckey, Hofstein, Mamlok-Naaman & Eilks, 2013; Clegg & Kolodner, 2014). Regarding this matter, Tuan & Chin (2000) demonstrated that concrete, perceptual and relevant science contents improve students' motivation when learning science. Therefore, it may be necessary to connect the contents studied in science class with the students' everyday life.

Concerning to the scientific language, several studies agree that language learning is an essential component when learning science in general (Wellington & Osborne, 2001; Brown & Ryoo, 2008) and chemistry in particular (Taber & García Franco, 2010; Song & Carheden, 2014; Childs, Markic & Ryan, 2015). As mentioned before, language is a great obstacle to most students when learning science (Wellington & Osborne, 2001). Concretely, the symbolic characters of chemical language may be considered as a significant contribution to its abstract nature (Beek & Louters, 1991;

Taskin & Bernholt, 2014). In addition, Pyburn, Pazicni, Benassi & Tappin (2013) have identified a relation between succeeding in completing a general chemistry course and the chemical language understanding.

Purpose

Under these circumstances, it is necessary to analyse the everyday situations related to scientific connotations and examine in detail the language used in those contexts. But to carry out this type of analysis, the first step is to leave the classroom, accompany the citizens in their day to day and analyse all the actions carried out from the perspective of the Teaching of Experimental Sciences (Pro & Ezquerria, 2004). By taking into account this information, we could establish references that indicate what kind of content should appear in the science curriculum and guide academic activities to suit the family situations for students. Once the scientific content is identified, it is possible to define the cognitive demands required by individuals in general and students in particular, for taking a well-founded decision. However, due to the immense amount of contexts, elements and contents where science is presented, we find it necessary to limit the scope of this work. Specifically, for this study, we have chosen one single context –the home science–, one element –the household cleaning products– and one written content –the information of the packaging and labelling of those products.

The reason why we chose this information is because household cleaning products are in every house and they are familiar to both citizens and students. Therefore, the use of this content can be used to work activities based on real life problems and connect the barrier between the real world and the classroom (Pine et al., 2006; Akçay, 2009). In addition, these products show interesting information for our aims on their labelling or packages, such as their chemical composition, safety instructions and conservation standards. These three aspects are regulated by the European Commission according the CE n° 1272/2008. In addition, most of this information is related to science. Regarding to this, a special Eurobarometer about consumer understanding of labels and the safe use of chemicals indicates that a sizeable number of respondents are unable to identify everyday household chemicals as chemical products. Moreover, 33% of respondents do not read the safety instructions before using a chemical product, and only 7% of those who read the instructions follow them fully. Under such circumstances, one of the conclusions of this special Eurobarometer is that further education regarding chemical products and their safety issues is needed (EC, 2011).

Taking these points into account, the objectives of the present study are:

1. To identify and categorize the scientific contents of labelling and packaging of household cleaning products.
2. To analyse the scientific language for each of these categories.
3. To connect and evaluate the importance of the language used in labels with the scientific literacy term and to evaluate some further social implications
4. To correlate and evaluate the importance of the language used in labels with scientific literacy and evaluate further some social implications (advertisement moral, critical purchasing, false public beliefs...)

Complementarily, some ideas of how to work these contents in class are proposed.

METHOD

a) Data collection

The information used for the study was obtained from labelling and packaging of household cleaning products of different countries (by alphabetical order): Belgium, Denmark, France, Germany, Italy, Poland, Portugal, Rumania, Spain and United

Kingdom. There were no significant differences between the information displayed in the same products of different brands and countries. Around 500 products from 10 different countries were analysed.

It has to be taken into account that the information collected from these labels was translated into English in order to complete the categorization. Thus, the information presented here does not reproduce them literally since they were written in different languages.

b) Data analysis

The methodology used to analyse the information presented in these labels was a content analysis of written text (Krippendorf, 2004). In order to isolate the scientific content of the units of information presented in these labels, we selected the information criteria to consider whether a content appearing on labels could be considered scientific or not. For that, we compare each of these contents with the 2015 PISA Science Framework (OECD, 2013). In this framework, scientific competence may be characterized as consisting of the next connected aspects: contexts, knowledge, competencies and attitudes. For this work, we took into account the different contexts and knowledge set by the OECD.

According to this framework, contexts are chosen according to the relevance to students' lives. Among all of them, we have selected those related to the topic of this study, the household cleaning products: *health (maintenance of health, accidents); natural resources (personal consumption of material and energy); environmental quality (use and disposal of materials and devices); and hazards (risk assessments of lifestyle choices)*. In relation to knowledge about the natural world -or knowledge of science-, PISA selects the items according to their relevance to real-life situations within the next categories: *physical systems (properties of matter, chemical changes of matter); living systems (human health); and technology systems (optimization, costs, benefits, risks)*.

Regarding to this, we were able to select a set of information units classified into several emerged and not exclusive categories: information referred to physical units, information referred to the composition of the product, information referred to procedures, information referred to security instructions, information referred to environmental specifications and information with advertising information.

In this way, information referred to physical units and composition is related to physical systems. All the contents referred to procedures and security instructions – *avoid contact with skin, do not ingest, do not mix with other products*, etc. - are related to the maintenance of the health and the avoiding of accidents. The information referred to environment specifications, as well as the one with advertising intention may be associated with the personal consumption of materials and energy, as well as to the environmentally friendly behaviour.

RESULTS and DISCUSSION

A detailed description and discussion of the information gathered from the labelling and packaging of household cleaning products of 10 different countries is described following the above-referred classification.

a) Information referred to physical units

This section shows objective information about features of the product. In the analysis process, it was observed that these units could be classified as predefined physical units and units of convenience.

The predefined physical units (Table I) are homogeneous and firmly established. Within this category, two types can be identified: the words themselves (density, concentration, conductivity...) and the symbols of the units. Each of these subcategories –words and symbols- has a different type of language. Each type of language used in science requires different cognitive demands on the individuals (Wellington & Osborne, 2001), and therefore it requires different cognitive demands on the consumers.

Table 1. *Information referred to predefined physical units*

Magnitude [unit]	Examples of some products
Mass [mg; g; kg]	Washing up liquid, dishwasher powder, detergent, bleach
Volume [L; mL; cm ³ ; pint]	Washing up liquid, bleach, floor cleaner, ironing water, glass and window cleaner
Density [g/mL; kg/L]	Washing up liquid, WC cleaner
Concentration [%; mg/L; g/mL]	Dishwasher, detergent
Temperature [°C; °F]	Detergent, dishwasher
Time [min; weeks; months; years]	Dishwasher, floor cleaner, window and glass cleaner
Distance [cm; mm]	Window and glass cleaner
Conductivity [μS]	Ironing water
pH [dimensionless]	WC cleaner, furniture cleaner
Acidity [%]	Plughole unblocked bathroom

Some of these units are well-established in our day-to-day (e.g. time or distance), so individuals should not find difficulties in understanding them. On the other hand, there are other units that, despite of being commonly used, may have dual meanings when used in colloquial or scientific language. According to several studies (e.g. Beck, McKeown & Kucan, 2002; Lee, Quinn, & Valdés, 2013) individuals tend to confuse both meanings. For instance, “heat” is used as synonymous of “high temperature” (Jasien & Oberem, 2002; Brookes & Etkina, 2015); “mass” is used as synonymous of “weight”; the word “acidity” is also used in colloquial language as sharp-tasting or sour, while in chemistry is used as the amount of acid present in a solution, often expressed in terms of pH. Usually, students do not identify the scientific meaning of this type of words, although the context where they are used is completely different. This highlights how language can be barrier to learning (Wellington & Osborne, 2001; Brown & Ryoo, 2008).

Another point to take into account is the use of different units to measure the same magnitude (e.g. volume in “cm³” and “ml”, concentration in “mg/L”, “g/100g” and “%”), which may cause difficulties and doubts in individuals. Likewise, it is also common to use the same units for different magnitudes (e.g. “%” to measure concentration and acidity; or “g/mL” to measure both concentration and density). Therefore, products that do not specify the magnitude that is being measured may be misleading.

Another kind of units has been identified and, because of their special features, we named them *units of convenience* (Table II). These terms are heterogeneous and seem to be created with the aim of measuring specific properties of the product by a unit created ad hoc. The language used in this type of units present a more familiar approach in order to facilitate its understanding to the consumers (e.g. measuring the volume of a detergent in *number of washings* instead of litres). It is worth mentioning that they do not follow pre-established rules. Therefore, the *units of convenience* try to bring scientific concepts and consumers closer.

Some of these *units of convenience* are widespread, for instance, measuring quantities in number of cups or tablets (“two caps, one tablet”) as commonly observed in detergents for dishwashers and washing machines. However, there are others that

are more whimsical, such as the level of the detergent concentration (“850 g normal detergent= 650 g concentrated detergent = 10 washes”).

Table 2. Information referred to units of convenience

Magnitude [unit]	Examples of some products
Quantity [cap, washes, tablet, bucket, washing]	Dishwasher, floor cleaner, bleach
Concentration [°F]	Dishwasher, detergent
New concentrated: 850 g = 650 g = 10 washes	Concentrated detergent
Water hardness [soft, medium and hard]	Detergent
2 caps (60 ml) in 5L of water	Floor cleaner

b) Information referred to composition of the product

Contrary to labelling of food, cosmetics and personal hygiene products, the legislation for household cleaning products do not oblige to show the complete composition, but only quantitative margins of the active substances, and other components of particular interest due to of their effect on environment or the presence of allergens (see section *information referred to security instructions*), among others.

When classifying this information (Table III), we realized that products used for different purposes have a very similar composition, since they share most of the active ingredients such as surfactants. Within the same kind of products, this similarity is further exacerbated, up to the point that some products of different brands have the same composition.

Most of these contents are technical words rarely used in colloquial language. Therefore, in order to succeed in the communication process, a connection should be established between the understanding of these words and the products' properties. However, sometimes it is not easy to find this link due to the fact that the information is not always expressed on the same basis. For instance, it is common to find in the same label components in different languages. Furthermore, in some products, information is displayed in both words and symbols (e.g. sodium hypochlorite or NaClO). This kind of language with two different components –nomenclature and symbols– is specific to chemistry. However, it is necessary to know that, even though they are expressed in a different way –either one language or another, either nomenclature or symbol-, they do mean the same.

Table 3. Information referred to the composition of the product

Chemical compound	Example of some products
Anionic surfactants	
Non-ionic surfactants	Kitchen cleaners; detergents, bathroom cleaners
Cationic surfactants	
Amphoteric surfactants	
Phosphonates	Detergents
Polycarboxilates	
Sodium hypochlorite (NaClO)	Bleach; bathroom cleaners; glass cleaner
Sodium hydroxide (NaOH)	Ammonia; bathroom cleaners; Specific cleaning products for
Potassium hydroxide (KOH)	the kitchen
Quaternary ammonium salts	Cleaning detergents
Ethanol; isopropyl alcohol	Detergents; fresheners and fragrances

Attending to the way of expressing the composition, chemical symbols own specific characteristics as they represent a high level of meanings. For instance, NaClO

formula “is telling us” that three different atoms –sodium, chlorine and oxygen– form this compound. In addition, the proportion of these atoms is 1:1:1. However, Arasasingham, Taagepera, Potter & Lonjers (2004) show that only fifty per cent of the students give right explanations about the meaning of subscripts in chemical formulae. Besides, if individuals had the proper knowledge, they would know that NaClO is an ionic compound that has the specific properties of these kinds of substances: soluble in water or other similar solvents, conducts electricity in liquid state or in solution, etc.

One fact that attracted our attention was that there are several products that advert in their packaging or labelling that it is free from one specific component, but this ingredient does appear in the composition. For instance, we have found detergents that are advertised as alcohol-free, but in the composition labelling does appear “ethanol” or other component that ends in –ol (representative of alcoholic compounds). Similarly, we have found floor cleaners announced as “bleach-free”, while they do have sodium hypochlorite; and hob cleaner “without detergent” but with surfactants within their components.

c) Information referred to procedures

Information showed in Table IV makes reference to procedures to be followed when using or storing the product. In this category, language is focused on instructions, which tends to be very simple and clear. However, they usually require the understanding of some conceptual contents, such as dilution, concentration or temperature.

Table 4. *Information referred to procedures*

Procedures	Examples of some products
Use a solution of 20 ml bleach to 5 litres of water	Bleach
50 ml per 10 Litres of water at 30°C	Floor cleaner
Dilute first	Floor cleaner, bleach
Spry sparingly from a distance of about 20 cm away from the surface	Window and glass cleaner
Tip the bottle to fill de neck	WC cleaner
No need to rinse	Floor cleaner
Outstanding at low temperatures	Detergent
Pour at least a third (100ml) of the bottle down the plughole	Plughole unblocked bathroom
Store in a cool dry place	Washing up liquid
Do not store below 4°C	Floor cleaner
Always reseal bottle and store upright	Floor cleaner, WC cleaner
Mop up spills immediately	Window and glass cleaner
On surfaces other than glass, mirror, tiles, test on an inconspicuous area first	Window and glass cleaner
Avoid use in aluminum/copper/brass	Floor cleaner
Not suitable for wool or silk	Bleach

In this section, guidelines are carefully given, indicating not only what to do (dilute first, for instance) but how to do it exactly (use a solution of 20 ml bleach to 5 litres of water, for example). Information is commonly displayed both textually, using words, and visually, using images represented by icons. According to this, Lemke (1998) indicates that in science, not only words are important but also the combination of other elements such as diagrams, images, tables, graphs, etc. However, these elements require different levels of understanding. For instance, tables and graphs (not very iconic but more symbolic) demand more training for the consumer in order to be interpreted. On the other hand, diagrams, photos and pictures, should require a lower level of training, due to their high iconic level. It could be considered

the more iconic character of the images, the lower level of education by the consumer. Therefore, since icons in labeling and packaging demand lower levels of understanding by consumer, they promote the comprehension of the instructions for use.

d) Information referred to security instructions

These contents refer to preventive measures, which indicate directly actions not to be done while using the product and measures to be taken in case of improper use of the product (Table V). The information presented here -legalised by CE No 1272/2008- is probably the most important content showed in the labelling. The misunderstanding of the instructions language may lead to a wrong follow-up of the protocol that can result in hazards such as intoxication or skin irritation, among others. Hence, the language of these contents must be understandable and easy to follow. However, there are some contents that are not clear and consequently, individuals may do not know how to act (e.g. “contains sodium hypochlorite and sodium hypochlorite solution”).

Apart from these indications, there are also some pictograms with warning symbols which provide information about possible damages that a specific substance can cause on health or the environment. These pictograms have been recently modified and unified according to the CE No 1272/2008 and the CLP (Classification, Labelling and Packaging of substances). As commented before, the combination of these elements facilitates the comprehension of the security instructions. Similarly, the two types of images –iconic and symbolic- demand different levels of training by consumers, being higher for symbols compared to icons, that are much more realistic.

Table 5. Information referred to security instructions

Security instructions	Examples of some products
Always read the label and product information before use	WC cleaner
Refer to manufacture’s washer instructions before use	Bleach
Keep out of the reach of children	Window and glass cleaner, WC cleaner
Do no ingest	All the products
People with sensitive or damaged skin should avoid prolonged contact with the product	Window and glass cleaner
Ventilate the room after use	Window and glass cleaner
Do not mix with bleach or any other household cleaners	WC cleaner
Contact with acids liberated toxic gas	Plughole unblocked bathroom
Wash hands after handling product	WC cleaner, detergent
May product allergic reaction	Floor cleaner
Causes serious eyes irritation	Floor cleaner, Spray cleaner
Always take care when using bleach on fabric	Bleach
Avoid breathing spray and point spay away from face	Window and glass cleaner
Do not breath spray	Window and glass cleaner
If swallowed, seek medical advice immediately and show this container or label	Plughole unblocked bathroom, WC cleaner
Contains sodium hypochlorite and sodium hypochlorite solution	Plughole unblocked bathroom
Causes severe skin burns and eye damage	Plughole unblocked bathroom
Do not breath fumes	Plughole unblocked bathroom
Wash skink thoroughly after handing	Plughole unblocked bathroom
If swallowed, seek medical advice immediately and show this container or label	WC cleaner, Window and glass cleaner
If product gets into eyes rinse thoroughly with water	Window and glass cleaner, WC cleaner

e) Information referred to environmental specifications

This group contains information related to the supposed environmental impact (Table 5). In this study we have just taken into account the information referred to the product itself, not to the packaging materials. Because of that, neither the terms nor the symbols about the extraction, fabrication or recycling process of the packaging are considered.

As can be observed in Table VI, most of these information are messages that suggest environmental benefits, but without substantiating argumentation. This practice -called “greenwash” or “greenwashing”- has been practiced by many companies in the last fifteen years, due to the fact that consumers have become more environmentally aware and companies take advantage of it (Delmas & Cuerel, 2011; Parguel, Benoit-Moreau & Russell, 2015).

Regarding to the household cleaning products, this practice is even more usual, as in these products the rate of “green” inventory growth is higher than in the overall products (101% versus 73%) (Terrachoice, 2010). The language used in this kind of messages includes words such as “eco-friendly”, “environmental-friendly”, “earth-friendly”, etc., which are ambiguous and may confuse rather than clarify. The use of this type of language is defined as “sin of vagueness”, which “*it is committed by every claim that is so poorly or broad that its real meaning is likely to be misunderstood by the consumer. All natural is an example. Arsenic, uranium, mercury and formaldehyde are all naturally occurring, and poisonous. All natural is not necessarily green*” (Terrachoice, 2010: 10). Producers and traders usually take advantage of the lack of knowledge of scientific language of a high number of target clients, by advertising their product through a fake image that improves the sales.

On the other hand, there are products that do show legitimate green certification, like the EU Ecolabel, which guarantees that products have complied with a set of criteria. This Ecolabel is an image which language facilitates the consumer identification of the products with this certification. Moreover, there are some messages with advises about how to save water, energy and CO₂. For instance, a specific diagram was created (commented in the information referred to security instructions category) in order to indicate that a substance is hazardous to the environment and causes aquatic toxicity.

Table 6. *Information referred to environmental specifications*

Environmental specifications	Examples of some products
Environmentally friendly	Plughole unblocked bathroom
Eco-chemistry	Detergent, WC cleaner
Environment safe	Detergent, dishwasher detergent
Use biocides safely	WC cleaner
Biodegradable	Bleach, detergent
Vegetal origin components	Detergent
Only vegetal origin surfactants	Detergent
No tested on animals	Detergent, bleach
It contains no phosphates	Dishwasher detergent, detergent
Non bio laundry powder contains among other ingredients	Detergent
Advices to save water, energy and CO ₂	Detergent, WC product
EU Ecolabel	All the products

f) Information with advertising intention

This category groups all contents that highlight a property of the product with the aim of differentiate it from other similar. The analysis of this information might be a whole report, however in this study we have only taken into consideration some general aspects. This information is shown with bright colours and big fonts in order to draw the consumers’

attention. Another common strategy used in marketing is to show a strong scientific and technological component (Belova, Rundgren & Eilks, 2015).

However, not all these contents use science in a proper way. Concretely, there are several examples where science was used incorrectly in advertising (Campanario, Moya & Otero, 2001). It is quite common to identify messages with *apparently scientific concepts that do not exist* (e.g. [brand] oxiaction; ultraconcentrated; polar strength; oxiaactive); *scientific concepts used in a wrong way* (e.g. power and pure; active power; biological active); *incorrect exaggerations* (e.g. kills E-coly, salmonella, MRSA, rotavirus, flu virus (H1N1); kills all germs; effective against salmonella, E-coli, listeria and MRSA; provides complete hygiene for your toilet; kills fly virus); information with *incomplete quantitative comparison* (e.g. better reach = better cleaning; less powder, more power; 99,99% germ kill; removes 90% of allergens); *erroneous and fallacious reasoning or arguments difficult to understand* (e.g. no enzymes or bleach; oxygenated additives; extra-bright polymers; very strong ammoniac; effective formula, natural power; more concentrated formula); and *advertisements that refer to science and technology as a source of authority* (e.g. recognising comfort, research into skin care. British skin foundation; dermatologically tested; stain lifting technology; new ultra power formula). This last practice is widespread and has been examined by several authors (Pitrelli, Manzoli & Montolli, 2006).

Another very common practice is the use of messages with negative connotations towards chemistry such as "the less chemical, the better" (Belova & Eilks, 2015). These authors remark the use of 'natural' in cosmetics advertising, which is extrapolated to other type of products including the household cleaning ones (e.g. less chemicals; contains pure orange oil; natural ingredients). Lastly, we have also found a wide amount of messages with an illegitimate use of the ecological and the green of the product (e.g. eco-chemistry; earth-friendly, etc). The use of this type of messages has been already discussed in the *Information referred to environmental specifications* category.

All these messages have an ambiguous, uncertain, indefinite and unclear language that may confuse individuals.

CONCLUSIONS

In the present work, the language used in the scientific contents of household cleaning products was analysed. Specifically, we detected simple words, symbols, formulas, pictograms, instructions with a more complete sentences, and advertising messages, some of them with lack of meaning. These facts evidence the great amount of science and technology references present in our everyday life, which highlights the necessity of providing a proper level of scientific literacy among individuals.

Since the purpose of language is communication, the message sent by the addresser should be understood by the addressee. In the particular case of producers and consumers of household cleaning product, this communication line can be impeded by several factors (labelling or packaging information removal, messages in foreign language, etc.), being language one of the main obstacles due to several reasons: (1) the meanings of the codes are not shared by addresser and addressee (do consumers understand chemical symbols?); (2) the message is not clear and concise (mistakes and exaggerations are widely used in advertising); (3) the addressee's previous ideas may obstruct or distort the information carried in the message ("heat" and "high temperature"; "mass" and "weight"; "acidity" and sharp-tasting or sour). Thus, these facts suggest different scenarios.

1. Producers and traders should be aware of the communication gap resulted from the different use of codes between addresser and addressee. The use of the right language might reduce this gap by linking the knowledge required and the knowledge own by

consumers (by using units of convenience, or by more educational advertisings, for instance). This approach improves the mutual understanding where both parts benefit, consumers know what and why they are making a purchase and producers can show specific advantages of their product.

2. The use of unclear and no-concise messages triggers the spread of false conceptions of science, such as being incomprehensible. This situation is used by traders and producers to maintain an asymmetry in their relation with consumers. According to the Principal-Agent theory (Eisenhardt, 1989), the consumer (principal) contracts a service, and producers and traders (agent) perform this service. But the consumer needs the agent's opinion (or a prior consultation) and this "technical advice" defines an information asymmetry, since the agent has business interests in the aid provided. Thus, they are not always interested in the consumers knowing certain information about their product. Obviously, the individual would achieve the required level to defend their rights.
3. The addressee's ideas may obstruct or distort the information of the message. The scientific background of citizens is heterogeneous and complex which affects the decision-making process in their daily life. In this scenario, public institutions should promote guiding initiatives with the aim of science comprehension among the population. However, this implies a deep analysis of all scientific contents present in society.

In order to achieve a proper level of scientific literacy, individuals should follow three steps: being able to read, to understand and to interpret the information. However, this is not always easy because of the technical of the language. In this context and taking into account several studies, Brown & Ryoo (2008) distinguish between "vernacular language" –used in everyday contexts–, and "non-vernacular language" –more specialised, used in more specific contexts (Lavob, 1972). According to these authors, both vernacular and non-vernacular languages are used "*based on the speaker's understanding of the intended audience*" (p.530). Since individuals in general and students in particular need to understand information on the labelling of household cleaning products, the language used should sound familiar to them; therefore, a conversion between non-vernacular to vernacular language is required. Or expressed in other terms, individuals need to code-switch the scientific language to the daily life one (Moje, Collazo, Carillo & Marx, 2001; Brown & Spang, 2008). If consumers do not understand the meaning of the labels, communication fails.

On the other hand, and regarding to the formal educational point of view, each type of language analysed implicates different affordances in science, which have to be taken into account in the learning and teaching process (Akaygun & Jones, 2014). In this sense, one of the elements to achieve the scientific literacy is to understand the scientific language, up to the point that understanding scientific language is considered one of the main issues when learning science. The success of reading comprehension is determined by the understanding of vocabulary in general (Moghadam, Zainal & Ghaderpour, 2012) and the scientific vocabulary in particular (Seah, Clarke & Hart, 2011; 2014). In short, the familiar and unfamiliar vocabulary determines the comprehension of texts or speeches (Hu & Nation, 2000; Schmitt, 2000; Stahl, 2003; Bahçivan, 2014). However, teachers rarely take into account their role as communicators (Borsese, 1994), which might explain previous failures in achieving scientific literacy. This lack of awareness makes them underestimate the deep differences between daily life and scientific language, which is vital to the teaching-learning processes (Pozo, Sanz, Gómez Crespo & Limón, 1991; Galagovsky, Bonán & Adúriz Bravo, 1998). Accordingly, here we sketch out some ideas or recommendations of how to work scientific languages in the classroom without going into details, leaving the specifics to teachers.

One of the tasks of science teachers is to teach students to read and understand scientific texts (Fang, 2006). However, it is complicated for readers to unpack the encoded information in these texts due to their lexically dense writing (Fang, 2005; Glass & Oliveira, 2014). In this sense, labelling and packaging information has short sentences and do not present a lexically dense writing style. Moreover, they are familiar to students. Taking this into account, we believe that labelling and packaging information may be considered as another scientific text to work with in science class. In fact, Calvo (2014) has already used the scientific content of labels (all kind of products, not only household cleaning ones) to teach chemistry nomenclature.

According to the analysis of the scientific language of household cleaning products performed in the present work, we consider that students should achieve the following goals (among others) in order to acquire scientific literacy: identify the units and the magnitudes; be aware of the industrial procedures for the fabrication of a production and its effects on the environment; connect the characteristics of a specific product with his or her necessities and establish the advantages and disadvantages of a product; identify different types of ingredients and explore their characteristics (Ngai, Sevian & Talanquer, 2014), recognize chemical compounds and encourage learning chemistry nomenclature, etc. Taking into account this last aspect, science and chemistry teachers should be aware of the difficulties and problems with the meaning of symbols (Marais & Jordaan, 2000). Thus, these authors suggest some recommendations of how to work these contents in class.

For that, activities should be oriented to solve theoretical questions (e.g. change of units, calculations of percentages and amount of substance); to carry out estimations (e.g. how many taps does this packet of detergent content?) and investigations (e.g. where do the products waste goes?); to develop scientific testing (e.g. what is the pH of this detergent) and critical thinking (Carvalho et. al, 2015). In addition, students' attitudes towards the effects of human activities on environment, modern purchasing and advertising, should also be worked through activities that evidences the necessity to analyse correctly situations using empirical data and avoiding personal opinions or public belief. Anyhow, these are only some ideas and orientations. Details and final execution will depend on each teacher.

REFERENCES

- Aikenhead, G. (2005). Science-based occupations and the science curriculum: Concepts of evidence. *Science Education*, 89(2), 242-275.
- Akaygun, S. & Jones, L.L., (2014). Words or Pictures: A comparison of written and pictorial explanations of physical and chemical equilibria, *International Journal of Science Education*, 36, 783-807.
- Akçay, B. (2009). Problem-based learning in science education. *Turkish Science Education*, 6(1), 26-36.
- Arasasingham, R.D., Taagepera, M., Potter, F. & Lonjers, S., (2004). Using knowledge space theory to assess student understanding of stoichiometry, *Journal of Chemical Education*, 81, 1517 – 1523.
- Bahçivan, E., (2014). Consistency among Turkish Students' Different Worlds: A Case Study Focusing on Responses to Science. *Journal of Turkish Science Education*, 11 (1), 101-114.
- Beek, K. & Louters, L., (1991). Chemical language skills: Investigating the deficit, *Journal of Chemical Education* 68, 389–392.
- Beck, I.L., McKeown, M.G. & Kucan, L., (2002). Bringing words to life: Robust vocabulary instruction. New York: Guilford Press.

- Belova, N. & Eilks, I., (2015). Learning with and about advertising in chemistry education with a lesson plan on natural cosmetics – a case study, *Chemical Education Research and Practice*, 16, 578.
- Belova, N., Rundgren, S. & Eilks, I., (2015). Advertising and science education: a multi-perspective review of the literature, *Studies in Science Education*, 51, 169-200.
- Borsese, A., (1994). Il problema della comunicazione linguistica a scuola: il linguaggio scientifico e chimico in particolare [The problem of linguistic communication in school: the language of science and chemistry in particular], *Enseñanza de las Ciencias*, 12, 333-337.
- Brookes, D.T. & Etkina, E., (2015). The Importance of Language in Students' Reasoning About Heat in Thermodynamic Processes, *International Journal of Science Education*, 37, 759-779.
- Brown, B.A. & Ryoo, K., (2008). Teaching science as a language: A “content-first” approach to science teaching, *Journal of Research in Science Teaching*, 45, 529-553.
- Brown, B.A. & Spang, E., (2008). Double talk: Synthesizing everyday and science language in the classroom, *Science Education*, 92, 708-732.
- Bybee, R. W. (2008). Scientific literacy, environmental issues, and PISA 2006: The 2008 Paul F-Brandwein lecture. *Journal of Science Education and Technology*, 17(6), 566–585.
- Calvo, M.A., (2014). Using Product Content Labels to Engage Students in Learning Chemical Nomenclature, *Journal of Chemical Education*, 91, 757-759.
- Campanario, J., Moya, A. & Otero, J., (2001). Invocaciones y usos inadecuados en la ciencia en la publicidad [Invocations and misuses science in advertising], *Enseñanza de las Ciencias*, 19, 45-56.
- Carvalho, C., Fíuza, E., Conboy, J., Fonseca, J., Santos, J., Gama, A. P., & Salema, M. H. (2015). Critical Thinking, Real Life Problems and Feedback in the Sciences Classroom. *Journal of Turkish Science Education*, 12 (2), 21-31.
- CE n° 1272/2008. Regulation (EC) No 1272/2008 of the European Parliament And Of The Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006.
- Childs, P.E., Markic, S. & Ryan, M.C., (2015). The Role of Language in the Teaching and Learning of Chemistry, in *Chemistry Education: Best Practices, Opportunities and Trends* (eds J. García-Martínez and E. Serrano-Torregrosa), Wiley-VCH Verlag GmbH and Co. KGaA, Weinheim, Germany.
- Clegg, T. y Kolodner, J. (2014). Scientizing and Cooking: Helping Middle-School Learners Develop Scientific Dispositions. *Science Education*, 98(1), 36–63.
- COSCE (2011). Informe ENCIENDE. Enseñanza de las Ciencias en la Didáctica Escolar para Edades Tempranas en España. [ENCIENDE Report. Science Teaching for Early Ages in Spain].
- DeBoer, G., (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform, *Journal of Research in Science Teaching*, 20, 582-601.
- Delmas, A. & Cuerel, V., (2011). The Drivers of Greenwashing. *California Management Review*.
- Duggan, S. & Gott, R., (2002). What sort of science education do we really need? *International Journal of Science Education*, 24, 661-679.
- EC (2006). Recommendation of The European Parliament and of the Council of 18 December 2006 On Key Competences For Lifelong Learning.

- EC (2011). Consumer understanding of labels and the safe use of chemicals. Special Eurobarometer 360. Brussels: European Commission.
- Eisenhardt, K., (1989). Agency theory: An assessment and review, *Academy of management review*, 14, 57-74.
- Fang, Z.H., (2005). Scientific literacy: A systemic functional linguistics perspective. *Science Education*, 89, 335–347.
- Fang, Z.H., (2006). The language demands of science reading in middle school, *International Journal of Science Education*, 28, 491–520.
- Galagovsky, L.R., Bonán, L. & Adúriz Bravo A., (1998). Problemas con el lenguaje científico en la escuela. Un análisis desde la observación de clases de ciencias naturales [Problems with scientific language in school. An analysis from the natural science classes observation], *Enseñanza de las Ciencias*, 16, 315-321
- Glass, R. & Oliveira, A. W., (2014). Science Language Accommodation in Elementary School Read-Alouds, *International Journal of Science Education*, 36, 577-609.
- Harlen, W., (2001). The assessment of scientific literacy in the OECD/PISA project, *Studies in Science Education*, 36, 79-104.
- Hodson, D., (2003). Time for action: Science education for an alternative future, *International Journal of Science Education*, 25, 645-670.
- Hu, M. & Nation, P., (2000). Unknown Vocabulary Density and Reading Comprehension, *Reading in a Foreign Language*, 13, 403-430.
- Jasien, P.G. & Oberem, G.E., (2002). Understanding of elementary concepts in heat and temperature among college students and K-12 teachers, *Journal of Chemical Education*, 79, 889-895.
- Kolstø, S., (2006). Patterns in Students' Argumentation Confronted with a Risk-focused Socio-scientific Issue, *International Journal of Science Education*, 28, 1689-1716.
- Krippendorff, K., (2004). *Content Analysis: An Introduction to Its Methodology*. 2nd edition, Thousand Oaks, CA: Sage.
- Labov, W., (1972). *Sociolinguistic Patterns*. Oxford, UK: Blackwell Press.
- Laugksch, R. C., (2000). Scientific literacy: A conceptual overview, *Science Education*, 84, 71-94.
- Lee, O., Quinn, H. & Valdés, G., (2013). Science and language for English language learners in relation to Next Generation Science Standards and with implications for Common Core State Standards for English language arts and mathematics, *Educational Researcher*, 42, 223-233.
- Lemke, J.L., (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin and R. Veel (Eds.), *Reading science*. London: Routledge.
- Lewis, J. & Leach, J., (2006). Discussion of Socio-scientific Issues: The role of science knowledge, *International Journal of Science Education*, 28, 1267-1287.
- Marais, P. & Jordaan, F., (2000). Are We Taking Symbolic Language for Granted? *Journal of Chemical Education*, 77, 1355-1367.
- Martin, M., Mullis, I., Foy, P. & Stanco, G., (2012). *TIMSS 2011. International Results in Science*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.
- Mercer, N., Dawes, R., Wegerif, R. & Sams, C., (2004). Reasoning as a scientist: ways of helping children to use language to learn science. *British Educational Research Journal*, 30, 367-385.
- Miller, J., (1983). Scientific Literacy: a Conceptual and Empirical Review, *Daedalus*, 112, 29- 48.

- Moghadam, S. H., Zainal, Z. & Ghaderpour, M., (2012). A Review on the Important Role of Vocabulary Knowledge in Reading Comprehension Performance, *Procedia Social and Behavioural Sciences*, 66, 555-563.
- Moje, E., Collazo, T., Carillo, R. & Marx, R. W., (2001). "Maestro, what is quality?": Examining competing discourses in project-based science, *Journal of Research in Science Teaching*, 38, 469-495.
- Ngai, C., Sevan, H. & Talanquer, V., (2014) What is ^[1]_{SEP} this Substance? What Makes it Different? Mapping Progression in Students' Assumptions about Chemical Identity, *International Journal of Science Education*, 36, 2438-2461.
- OECD (2007). PISA 2006 Science Competences for Tomorrow's World. Executive Summary.
- OECD (2013). PISA 2015. Draft Science Framework.
- Osborne, J., & Collins, S. (2001). Pupils' views of the role and value of the science curriculum: a focus-group study. *International Journal of Science Education*, 23(5), 441-467.
- Osborne, J., & Dillon, J. (2008). Science education in Europe: Critical reflections. London: Nuffield Foundation.
- Parguel, B., Benoit-Moreau, F. & Russell, C., (2015). Can evoking nature in advertising mislead consumers? The power of 'executional greenwashing', *International Journal of Advertising*, 34, 107-134.
- Pine, J., Aschbacher P., Roth E., Jones M., McPhee C., Martin, C., Phelps, S., Kyle, T., & Foley, B. (2006). Fifth graders' science inquiry abilities: A comparative study of students in hands-on and textbook curricula. *Journal of Research in Science Teaching*, 43(5), 467-484.
- Pitrelli, N., Manzoli, F. & Montolli, B. (2006). Science in advertising: uses and consumptions in the Italian press, *Public Understanding of Science*, 15, 207-220.
- Pozo, J.A. & Gómez Crespo, M.A., (1998). Aprender y enseñar ciencia [Learning and teaching science] Morata: Madrid.
- Pozo, J.A., Sanz, J.A., Gómez Crespo, M.A. & Limón, M., (1991). Las ideas de los alumnos sobre la ciencia, una interpretación desde la psicología cognitiva [Students' ideas about science, an interpretation from the cognitive psychology], *Enseñanza de las Ciencias*, 9, 83-94.
- Pro, A., y Ezquerria, A. (2004). La enseñanza de la Física: Problemas clásicos que necesitan respuestas innovadoras [The teaching of physics: classical problems that need innovative responses]. *Alambique: Didáctica de las Ciencias Experimentales*, 41, 54-67.
- Pyburn D. T., Pazicni S., Benassi V. A. & Tappin E. E., (2013), Assessing the relation between language comprehension and performance in general chemistry, *Chemical Education Research and Practice*, 14, 524-541.
- Schmitt, N. (2000). Vocabulary in language teaching. UK: Cambridge University Press.
- Seah, L.H., Clarke, D.J. & Hart, C.E., (2011). Understanding students' language use about expansion through analyzing their lexicogrammatical resources, *Science Education*, 95, 852-876.
- Seah, L.H., Clarke, D.J. & Hart, C.E., (2014). Understanding the Language Demands on Science Students from an Integrated Science and Language Perspective, *International Journal of Science Education*, 36, 952-973.
- Song Y. & Carheden S., (2014). Dual meaning vocabulary (DMV) words in learning chemistry, *Chemical Education Research and Practice*, 15, 128-141.
- Stahl, S. A., (2003). Vocabulary and readability: How knowing word meanings affects comprehension. *Topics in Language Disorders*, 23, 241-247.

- Stuckey, M., Hofstein, A., Mamlok-Naaman, R. y Eilks, I. (2013). The meaning of 'relevance' in science education and its implications for the science curriculum. *Studies in Science Education*, 49(1), 1–34.
- Taber K.S. & García Franco A., (2010). Learning processes in chemistry: drawing upon cognitive resources to learn about the particulate structure of matter, *Journal of the Learning Sciences*, 19, 99–142.
- Taskin, V. & Bernholt, S., (2014) Students' Understanding of Chemical Formulae: A review of empirical research, *International Journal of Science Education*, 36, 157-185.
- Terrachoice, (2010). The sins of greenwashing. Home and family edition. A report on environmental claims made in the North American consumer market.
- Tuan, H.L, Chin, C.C. & Shieh, S., (2005). The development of a questionnaire to measure students' motivation towards science learning, *International Journal of Science Education*, 27, 639–654.
- Tuan, H.L. & Chin, C.C., (2000). Promoting junior high school students' motivation toward physical science learning (III). Report for Taiwan National Research Council (NSC 89-2511- S018-030).
- UNESCO (United Nations Educational, Scientific and Cultural Organization) (2008). *Science Education Policy-Making: Eleven emerging issues*. UNESCO:Paris.
- Wellington, J.J. & Osborne, J., (2001). Language and literacy in science education. Philadelphia, PA: Open University Press.

Biology Education Research in Turkey: Trends from 1989 to 2015

Mustafa DERMAN¹ 

¹Dr., Atatürk University, Kâzım Karabekir Education Faculty, Erzurum- TURKEY

Received: 09.05.2016

Revised: 20.10.2016

Accepted: 01.12.2016

The original language of article is English (v.14, n.1, March 2017, pp. 89-109, DOI: 10.12973/tused.10192a)

ABSTRACT

This study was conducted to reveal out which biology subjects were preferred the most in Turkey in postgraduate theses; and to make comparisons of the university biology education, the types of publications, the research methods and investigated subjects. It was also aimed to find out the intensity of biology subjects dealt with in primary school and secondary school curriculum with regards to course hours on the basis of both the grades and the biology subjects. The qualitative research method was used in this study. The content analysis was used in data analysis. The result of study showed that; before 2000, MS theses were mostly done and PhD theses were fewer. Secondary school level was mostly preferred as for the research group and the “Quantitative” method was mostly preferred as for the research method. In addition, it was revealed out that; the environment, cell and variety of the living beings were among the topics that were studied the most in researches while the tissues, evolution, systems and energy were studied the least.

Keywords: Biology, education, content analysis, curriculum

INTRODUCTION

Interests, objectives, attitudes, concerns and motivation have great role in academic success and learning. These factors are directly interrelated and they have a significant impact on the behaviors of the individuals. In science education, there is a relation between the success of individuals and their interest in this area. The interest in science has a positive impact on the individual's success in this area (Hong, Shim and Chang, 1998; Prokop, Prokop and Tunnicliffe, 2007; Usak et al., 2009). Many students find science boring and difficult. However, this is not true for all branches of science areas. Particularly female students are known to have more interest in biology compared to others. Actually, the interest shown in biology is more than the others in general (Uitto, Juuti, Lavonen and Meisalo, 2006; Prokop, Prokop and Tunnicliffe, 2007). Biology has a unique place in that, it includes both laboratory and field studies (Prokop, Prokop and Tunnicliffe, 2007). Research methods utilized by the scientists are commonly utilized in biology researches. Observation, comparison, experimental studies and models can be listed among these (Krell and Krüger, 2015). Besides teaching the scientific issues, the science curriculum shall encourage the student to think like



a scientist (Roberts 2001). In science education and in other science areas (Prokop, Tuncer and Chuda, 2007; Harurluoğlu and Kaya, 2009), the teacher and the activities they perform in the class has the greatest role in bringing the students to the desired level and raising them as qualified individuals. This is because, the teacher is the one who knows how the students learn and what is the best way of learning (Prokop, Tuncer and Chuda, 2007). All the same; the methods utilized by the teacher in class, his/her ability to organize the activities and ever improve himself/herself also have a significant role in achieving the attainment of the lesson (Kaya and Gürbüz, 2002; Altunoğlu and Atav, 2005). Particularly in recent years, natural sciences have developed significantly as distinct from social sciences. This has brought about technological advancement and new information and thus, the need to teach the new information to the individuals through science education has arisen (Reiss, 2006). Teachers have the greatest role in achieving the objectives of this information and in the operation of the system. This is because, teachers are the ones who play an active role in bringing the students to the desired level and enabling the objectives to be completely achieved (Altunoğlu and Atav, 2005; Yılmaz and Çimen, 2008; Delibaş and Babadoğan, 2009). Education is expected to be given to individuals in classroom environment and through the activities utilized by the teacher (Uitto, Juuti, Lavonen and Meisalo, 2006). In order to achieve this, the teachers need to possess the essential pedagogical content knowledge and skills and there must be sufficient level of materials and activities for utilization (Kahyaoğlu and Yavuzer, 2004; Saka and Akdeniz, 2006). Classroom materials help the teachers to enable materialization of the issues and to contribute to the students' understanding of the lesson. Necessary methods and equipment shall be made available to teachers (Akaydın, Güler and Mülâyim, 2000; Köseoğlu and Soran, 2005). Particularly in science education, providing the students with sufficient level of experimental activities and practice would enable the students to be trained as a scientist and contribute much to the interest and success in science (Zhai, Jocz and Tan, 2014). This contribution shall not be limited to education but shall also enable the individuals to gain success in their future business lives (Reiss, 2006).

Literature review

Review studies made in science education provide great benefits for researchers. These studies both give ideas to researchers about the tendencies in science education and they guide them for their future researches (Tsai and Wen, 2005; Lee, Wu and Tsai, 2009). This is not only useful for the researchers but also for the persons and entities that have role in decision-making processes of the education system and it will help them improve the quality of education (Chang, Chang and Tseng, 2010). When previous studies are viewed within the scope of the research, it is observed that there are comprehensive researches on the area of science education; researches on science published on various journals between 1998 and 2002 (Tsai and Wen, 2005), researches on science education published on various journals between 2003 and 2007 (Lee, Wu and Tsai, 2009), between 2008 and 2012 (Lin, Lin and Tsai, 2014), 633 researches on biology education published on journals in Turkey on national and international journals (Gül and Sözbilir, 2015), 46 researches on physics education published between 2004 and 2011 (Önder et al., 2013), 53 researches on the environmental issues at primary education level, performed between 1997 and 2007 (Erdoğan, Marcinkowski and Ok, 2009), 314 researches on science education performed between 2005 and 2006 (Erdem, 2011) and Physics (66), Chemistry (49) and Biology (108) theses published on the areas of science education between 1990 and 2009 (Doğru, Gençosman, Ataalkın and Şeker, 2012). These researches show that; most of the researches consist of experimental studies; the amount of case, theoretical and review studies is few; mainly the environmental and ecological issues are handled in biology studies; the issues of learning, teaching and

attitudes are among the ones that are mostly studied; “program”, “misconception” and “material development” are the issues that are studied most in physics, chemistry and biology theses. (Tsai and Wen, 2005; Gül and Sözbilir, 2015; Erdoğan, Marcinkowski and Ok, 2009; Lee, Wu and Tsai, 2009; Erdem, 2011; Doğru, Gençosman, Ataalkın and Şeker, 2012; Önder et al., 2013). A new study was published by Lin, Lin and Tsai (2014) showed that in recent years “*argumentation, inquiry-based learning, scientific modeling, students’ learning, science teaching, and students’ conceptual learning*” were mostly researched by science education scholars. As can be seen in the detailed researches made in this area, there are two research-type studies in Turkey; Gül and Sözbilir (2015) studied the articles at national and international level whereas Doğru, Gençosman, Ataalkın and Şeker, 2012) made research on physics, chemistry and biology studies that took place in science education and as biology studies were taken into consideration, few (108) sources were reached at. Furthermore, in most of the review studies mentioned above, the methods and analysis techniques utilized in case studies were dealt with. On the other hand, the extent the topics making up a discipline are included in researches and whether these are in conformity with the curriculum were generally neglected. Due to the lack of a comparative research related to the universities teaching biology, this research was made to reveal out the extent master’s and doctoral degree theses done in Turkish universities teaching biology covered the topics of biology. Our comparative research aimed to reveal out the distribution of biology subjects matters concerning the studies in universities and the areas studies. Comparisons with the curriculum were made to help the researchers and the decision maker managers in the education system. In parallel to these, the following issues were inquired;

What was the distribution of theses among the universities?

What were trends in biology theses in term of publication type, sample, research type?

What elements were frequently investigated in term of university, publication type, sample, research method?

What subjects and subject matters were frequently investigated in term of university, publication type, sample, research method?

METHODS

The qualitative research method was used in this study. The qualitative research method basically makes use of the target text to create the category system, it organizes these to reveal out inter-categorical relations and features (McMillan and Schumacher, 2010). The content analysis was used in data analysis. The inductive and deductive approaches can be used together in content analysis. Yet, the inductive approach is mostly preferred. As the texts are read, the concepts and categories are formed. The available category system can also be used in content analysis (Elo and Kyngas, 2008). The codes were used for this study including publication type, sample, publication year and research method. The codes that listed above were used in many study for review the science literature (Tsai and Wen, 2005; Lee, Wu and Tsai, 2009; Doğru, Gençosman, Ataalkın and Şeker, 2012; Gül ve Sözbilir, 2015; Ormanci, Cepni, Deveci and Aydin, 2015). Besides this, some studies included highly-cited papers, number of authors and authors nationality etc. (Tsai and Wen, 2005; Lee, Wu and Tsai, 2009; Chang, Chang, and Tseng, 2010; Gül ve Sözbilir, 2015; Ormanci, Cepni, Deveci and Aydin, 2015). In the study, Content Analysis Form (CAF) (Appendix-1) was generated according to the purpose of the research. The content analysis form includes; the university the study was made in, publication type (MS and PhD), publishing year, sample that the levels the study was intended for (Primary School-PS, Secondary School-SS, High Education-HE and Others), the research method (Qualitative, Quantitative and Mixed), elements that the topics studied

(consists of 13 concepts) and the biology topic chosen (8 main subjects and 34 subjects matters related to these). Biology topics were determined from primary and secondary science curriculum (MEB 2013a and MEB 2013b). As can be seen in Table 1;

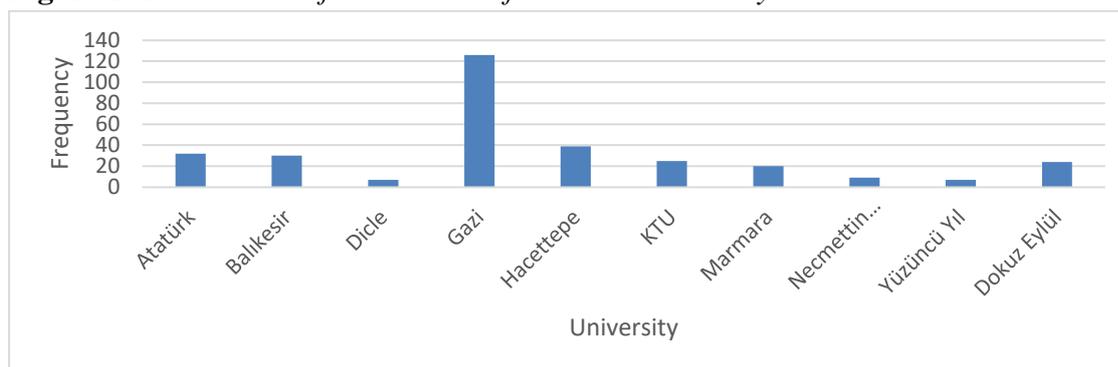
Table 1. *Biology subjects and subjects matters*

Biology:Life Sciences (BLS)	Cell (C)	Genetics and Heredity (GH)	Environmental Education (EE)
-The Nature of Scientific Knowledge and Biology (NSKB) - Common Features of Living Things (CFLT) -Compounds of Living Things (CLT)	- Substance Transport (ST) - Diversity and Classification of Living Things (DCLT) - Kingdom and Features (KF) - Cell Division (CD) -Cell (Ce)	- Reproduction, Growth and Development (RGH) - Modern Genetics and Biotechnology (MGB) - From Gene to Protein (FGP) -Basic Principle of Heredity (BPH)	- Current Environmental Issues (CEI) - Ecosystem Ecology (EEc) - Biomes (Bs) - Community Ecology (CEc) - Population Ecology (PE)
Evolution (E)	Tissues (Ts)	Systems (Ss)	Energy (Egy)
-Behavior (Be) -The Origin of Life (OL) - Evolution (Ev)	- Plant tissues and Plant Construction (PTPC) - Hormones and Homeostasis (HH) - Substance Transport in Plants (STinPs) -Animal Tissues (AT)	- Nervous System (NS) - Digestive System (DS) - Respiratory System (RS) - Urinary System (US) - Circulatory System (CS) - Muscular and Skeletal Systems (MSSs)	- Energy and Vitality (EV) - Photosynthesis (Ph) - Cellular Respiration (CR) -Chemosynthesis (Ch)

The Higher Education Council National Theses Centre was referred to as the source of data. 319 theses were reached within the context of graduate school of natural and applied sciences and graduate school of educational sciences at universities conducting MS and doctoral degree studies in biology education. No key words were used for the relevant research. The main reason for that was to exclude the theses on science education and to avoid re-counting of the study. In the research, institutes of the universities were chosen one by one through the National Theses Centre and all the studies conducted between 1989 and 2015 were aimed to reach at in chronological order. For the analysis of data SPSS 16 and Microsoft Excel were used. Findings of research were presented as both figure and table. Figures were only showed with frequency. But tables were showed percentage and frequency together. It done to express data more clearly.

FINDINGS

Findings of the research are presented in percentage-frequency and graphics depending on the universities, type of publication, year of publishing, the research method, the issues investigated elements and the research topic.

Figure 1. Distribution of the amount of thesis on university basis

As can be seen in Figure 1, “Gazi” University possesses the greatest amount of master’s degree theses. “Atatürk” (10.3%), “Balikesir” (9.4%), “Gazi” (39.5%), “Hacettepe” (12.2%), “KTU” (7.8%), “Marmara” (6.3%) and “Dokuz Eylül” (7.5%) universities are among the universities that have conducted 20 researches and more whereas “Dicle” (2.3%), “Necmettin Erbakan” (2.8%) and “Yüzüncü Yıl” (2.3%) universities have conducted fewer amount of researches. It can be stated that; studies on biology education conducted in universities remain at low level. In certain universities (Yüzüncü Yıl, Necmettin Erbakan and Dicle), this amount is even lower.

Table 2. Distribution of the research elements on university basis

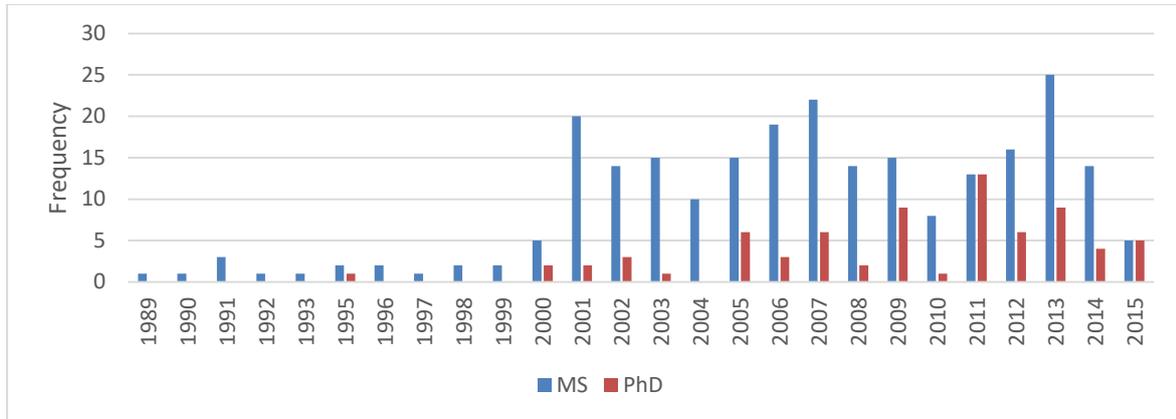
University	MS	PhD	PS	SS	HE	Others	Qualitative	Quantitative	Mixed
Atatürk	6(16)	22(16)	13(5)	11(17)	11(5)	8 (5)	10(4)	9(21)	11(7)
Balikesir	10(24)	8(6)	42(16)	6(10)	6(2)	3(2)	17(7)	7(16)	27(17)
Dicle	2(4)	4(3)	0	3(4)	3(1)	3(2)	0	3(6)	2(1)
Gazi	41(101)	34(25)	24(9)	37(58)	37(32)	45(27)	22(9)	44(99)	29(18)
Hacettepe	12(30)	12(9)	0	14(22)	14(7)	17(10)	20(8)	10(24)	11(7)
KTU	9(22)	4(3)	13(5)	8(13)	8(2)	8(5)	8(3)	7(16)	9(6)
Marmara	6(16)	6(4)	5(2)	5(8)	5(6)	7(4)	5(2)	7(16)	3(2)
Necmettin Erbakan	4(9)	0	0	4(6)	4(3)	0	5(2)	3(6)	2(1)
Yüzüncü Yıl	3(7)	0	0	4(6)	0	2(1)	13(5)	1(2)	0
Dokuz Eylül	7(17)	10(7)	3(1)	8(13)	8(6)	7(4)	0	9(20)	6(4)

*Frequency (f) was in parentheses. Percent (%) was outside of parentheses.

As can be seen in Table 2; in all of the universities, the share of master’s degree theses is higher in the distribution of post graduate theses. The greatest share belongs to Gazi University in terms of both “MS” (41%) and “PhD” (34%). The lowest share belongs to “Yüzüncü Yıl” University in terms of “MS” theses and to “Necmettin Erbakan” (0%) and “Yüzüncü Yıl” (0%) universities in terms of “PhD” theses.

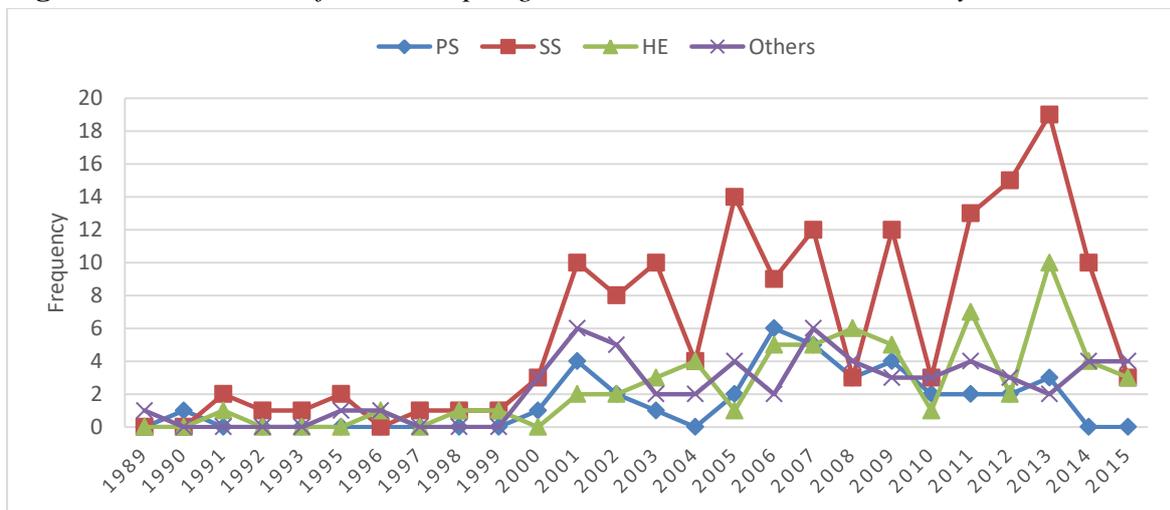
In terms of the levels the studies conducted in universities were intended for; “Atatürk”, “Balikesir”, “Gazi”, “KTU”, “Marmara” and “Dokuz Eylül” conducted studies intended for all the levels. The highest values belong to “Gazi” University in “SS” (37%), “HE” (37%) and “Others” categories and to “Balikesir” University in “PS” (42%) category. The lowest values belong to “Dicle”, “Hacettepe”, “Necmettin Erbakan” and “Yüzüncü Yıl” in “PS” category; to “Yüzüncü Yıl” in “HE” category and to “Necmettin Erbakan” University in “Others” category.

As for the categories related to the research method; the highest values belong to Gazi in “Qualitative” (22%), “Quantitative” (44%) and “Mixed” (29%) categories; the lowest values belong to “Dicle” (0%) and Dokuz Eylül” (0%) in “Qualitative” category; and to “Yüzüncü Yıl” in “Quantitative” (1%) and “Mixed” (0%) categories.

Figure 2. Distribution of theses on year basis

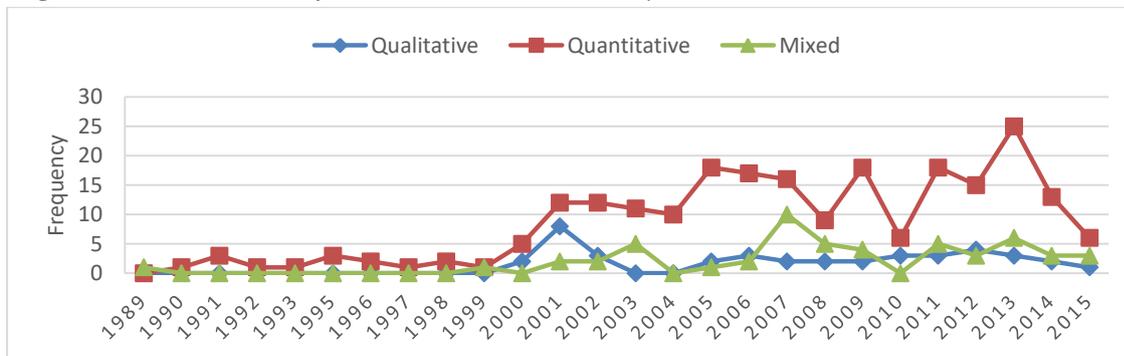
The amount of publications as for the postgraduate theses done in universities on biology education is rather low in before 2000. There is a rapid increase as of the year 2000 and several postgraduate theses have been published since then. The amount of “PhD” studies have increased, as well. The greatest amount of publications belongs to year 2013. “MS” studies were mainly conducted before 1999 and “PhD” studies were only conducted in 1995. From 1989 to 2015, only in 1994 no study on Biology education was conducted in universities. There is a decrease in the amount of postgraduate theses after 2013. There has been significant increase in the amount of postgraduate theses after 2000. Similarly; while there was increase in the amount of doctoral theses, decreases were observed in 2003 and 2010 (Figure 2).

While after 1999 there was increase in the amount of postgraduate theses intended for all level, decreases were observed in some years. As for the primary schools, there is no publication after 2013. Before 2000, the amount of studies was low for all levels. In terms of the research target group, the secondary education level can be stated to be preferred most by the researches. “SS” and “Others” are observed to be among the most preferred research target groups after 2000 (Figure 3).

Figure 3. Distribution of the levels postgraduate theses were intended on year basis

The “Quantitative” method is preferred more in universities. Though the utilization of “Qualitative” and “Mixed” research methods showed increase in time, they were preferred less by the universities in general. Whereas the researchers utilized the “Quantitative” method before 2000, the “Qualitative” and “Mixed” methods also started to be preferred after 2000 (Figure 4).

Figure 4. *Distribution of the research method on year basis*



Though there has been increase in the types of research methods utilized after 2000, the greatest increase has been observed in the quantitative methods. Besides the quantitative methods, the researchers have begun to prefer the qualitative and mixed methods, as well. As for the research methods, the highest values belong to the “Quantitative” method in 2013, the “Qualitative” method in 2001 and the “Mixed” method in 2007. On the other hand; it is possible to say that the “Qualitative” method has remained at a certain level after 2005 (Figure 4).

Figure 5. *General distribution of the investigated elements*

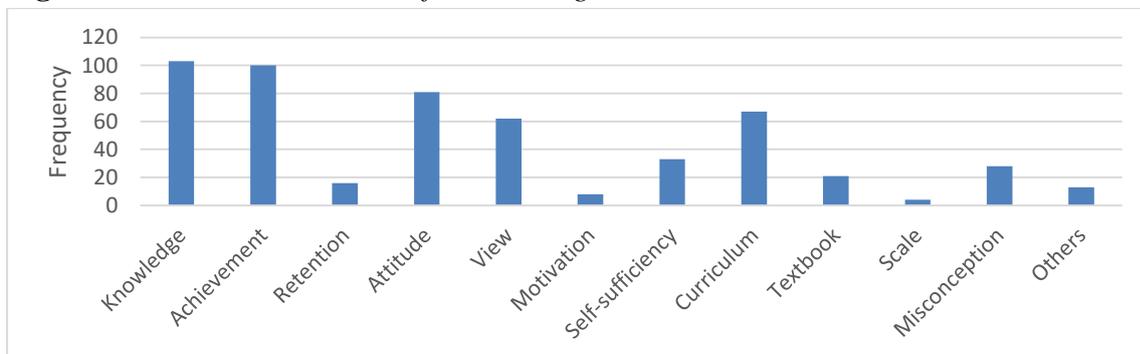
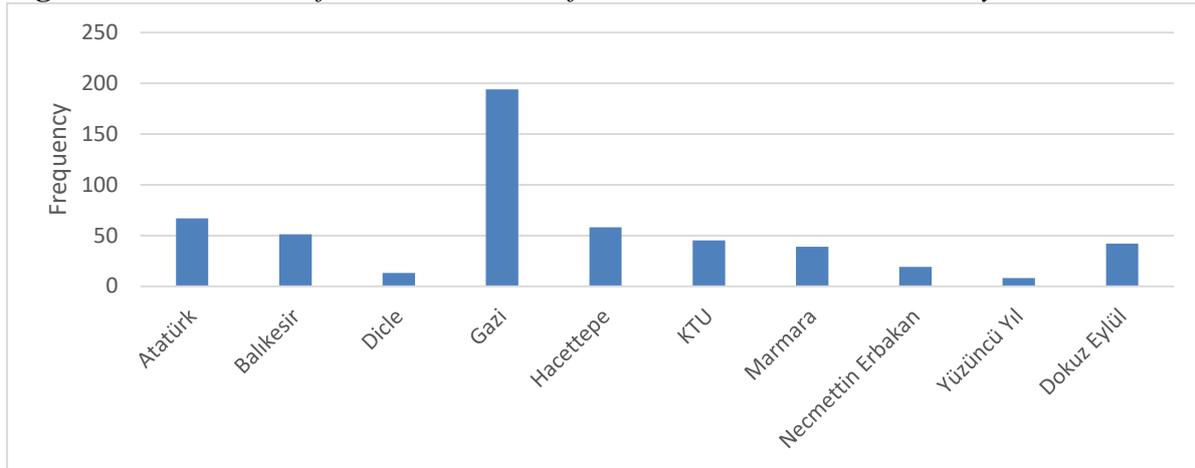


Figure 5 shows that; “Knowledge” (19.2%), “Achievement” (18.7%), “Attitude” (15.1%), “View” (11.7%), “Self-Sufficiency” (6.2%), “Curriculum” (12.5%), “Textbook” (3.9%) and “Misconception” (5.2%) elements were investigated the most. Categories of “Retention” (3%), “Motivation” (1.5%), “Scale” (0.8%) and “Others” (2.4%) were investigated less. It can be stated that, “Knowledge” has been the category focused on most and “Scale” has been the category focused on the least by the universities in the postgraduate studies.

Figure 6. Distribution of the total amount of research elements on university basis

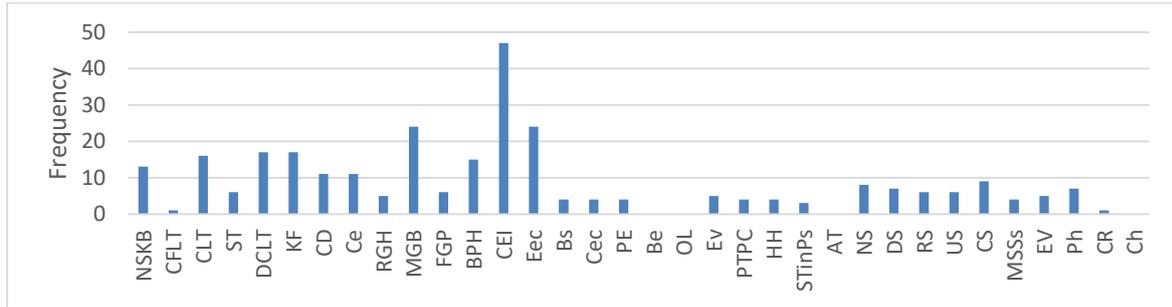
With regards to the elements consisting of 12 categories, the highest value belongs to “Gazi” University. Apart from “Gazi” university, there are two universities (Atatürk and Hacettepe) that take a value of 50 and higher. “Yüzüncü Yıl” University can be stated to take the lowest value with regards to the total number of elements covered.

As shown in Table 2; “Dicle”, “KTU”, “Marmara” and “Yüzüncü Yıl” universities are among the ones that have been unable to take the first and second highest values. The first or second highest values with regards to the elements investigated, are observed to belong to “Gazi” University. In “Atatürk” University, “Retention”; in “Balıkesir” University, “Misconception” and in “Hacettepe” University, “Textbook” has been among the elements with the highest values.

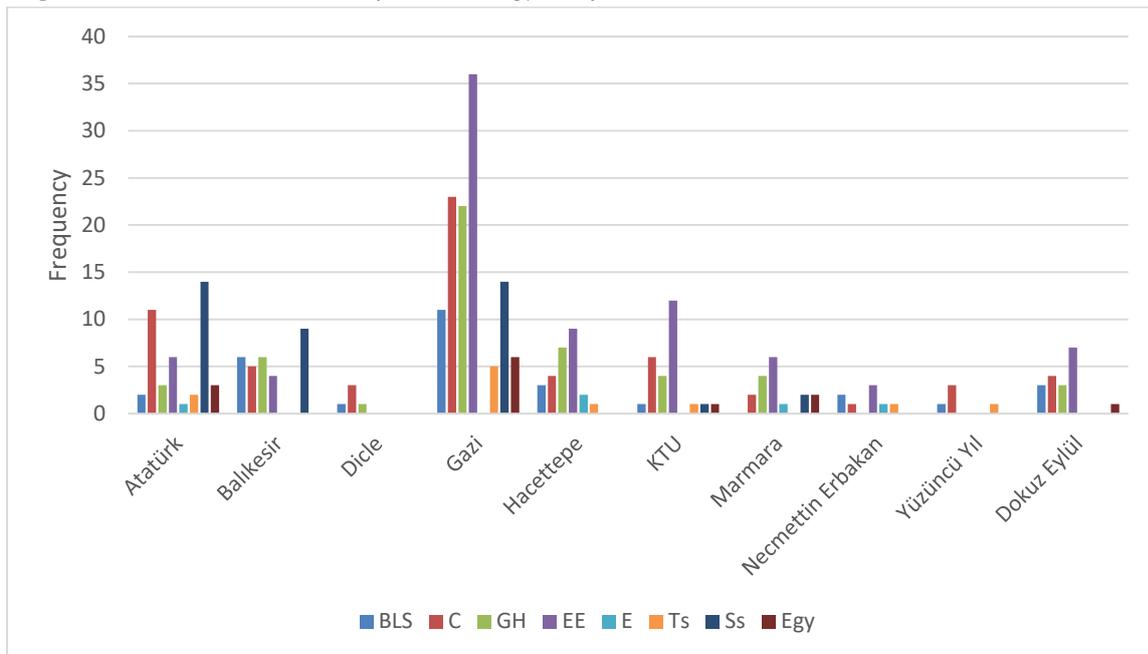
Table 3. Distribution of the research elements on university basis

	Atatürk	Balıkesir	Dicle	Gazi	Hacettepe	KTU	Marmara	N.Erb.	Y. Yıl	D. Eylül	Total
Knowledge	13.6(14)*	13.6(14)*	1.9(2)	30.1(31)*	9.7(10)	10.7(11)	10.7(11)	4.9(5)	1(1)	3.9(4)	103
Achievement	16(16)**	9(9)	2(2)	46(46)*	4(4)	7(7)	5(5)	0	0	11(11)	100
Retention	50(8)*	6.3(1)	0	31.3(5)**	6.3(1)	0	6.3(1)	0	0	0	16
Attitude	9.9(8)	12.3(10)	4.9(4)	34.6(28)*	4.9(4)	6.2(5)	6.2(5)	6.2(5)	1.2(1)	13.6(11)**	81
View	11.3(7)**	8.1(5)	1.6(1)	40.3(25)*	11.3(7)**	8.1(5)	8.1(5)	3.2(2)	1.6(1)	6.5(4)	62
Motivation	0	12.5(1)	12.5(1)	37.5(3)*	25(2)**	0	0	12.5(1)	0	0	8
Self-sufficiency	6.1(2)	0	9.1(3)	54.5(18)*	6.1(2)	3(1)	9.1(3)	0	0	12.1(4)**	33
Curriculum	9(6)	6(4)	0	28.4(19)*	25.4(17)**	10.4(7)	6(4)	3(2)	6(4)	6(4)	67
Textbook	9.5(2)	0	0	19(4)**	38.1(8)*	9.5(2)	14.3(3)	4.8(1)	4.8(1)	0	21
Scale	0	0	0	75(3)*	0	0	0	25(1)*	0	0	4
Misconception	10.7(3)	25(7)*	0	21.4(6)**	7.1(2)	17.9(5)	7.1(2)	7.1(2)	0	3.6(1)	28
Others	7.7(1)	0	0	46.2(6)*	7.7(1)	15.4(2)	0	0	0	23.1(3)**	13
Total											536***

*The highest value with regards to the categories. ** The second highest value. *** This value appeared as more than one element was investigated in the same research.

Figure 7. General distribution of the biology topics studied in researches

“CEI” (16%) is observed to be the topic that is most preferred by the researchers. “Be”, “OL”, “AT” and “Ch” on the other hand, are not preferred at all. Whereas the topics from “NSKB to Eec” are particularly preferred, the topics from “Bs to Ch” are preferred less. As for the 34 sub topics determined in the research; topics related to Power transformations, systems and tissues are observed to be at lower frequency ranges. As for the biology topics dealt with in the studies, it is possible to say that there has been a greater focus on certain topics.

Figure 8. Total distribution of the biology subjects studied in researches

With regards to the biology topics that are dealt with in 8 categories, “Atatürk” University is the only one that conducted study on all the topics. In terms of the topics preferred most, it is observed that; the topics “BLS”, “C”, “GH”, “EE”, “Ts” and “Egy” are mostly preferred in “Gazi” University; “SS” in “Gazi” and “Atatürk”; “E” in “Atatürk”. In terms of the total number of topics, the lowest figures belong to “Dicle” and “Yüzüncü Yıl” universities. It is also observed that; “EE” topic has attracted the greatest attention in all universities except for “Atatürk”. “Ss” topic has been mainly studied in “Balıkesir”, “Gazi” and “Atatürk” universities.

With regards to the values related to biology topics, the highest values belong to “Gazi” in “NSKB” and “CLT”; “Dokuz Eylül” in “CFLT”, “Gazi” in “ST”, “DCLT”, “KF”, and “Ce”; “Gazi” and “Atatürk” in “Ce”; “KTU” in “RGH”; “Gazi” in “MGB”; “Gazi” and “Hacettepe” in “FGP”; “Gazi” and “Marmara” in “BPH”; “Gazi” in “CEI”, “EEc”, “Bs” and

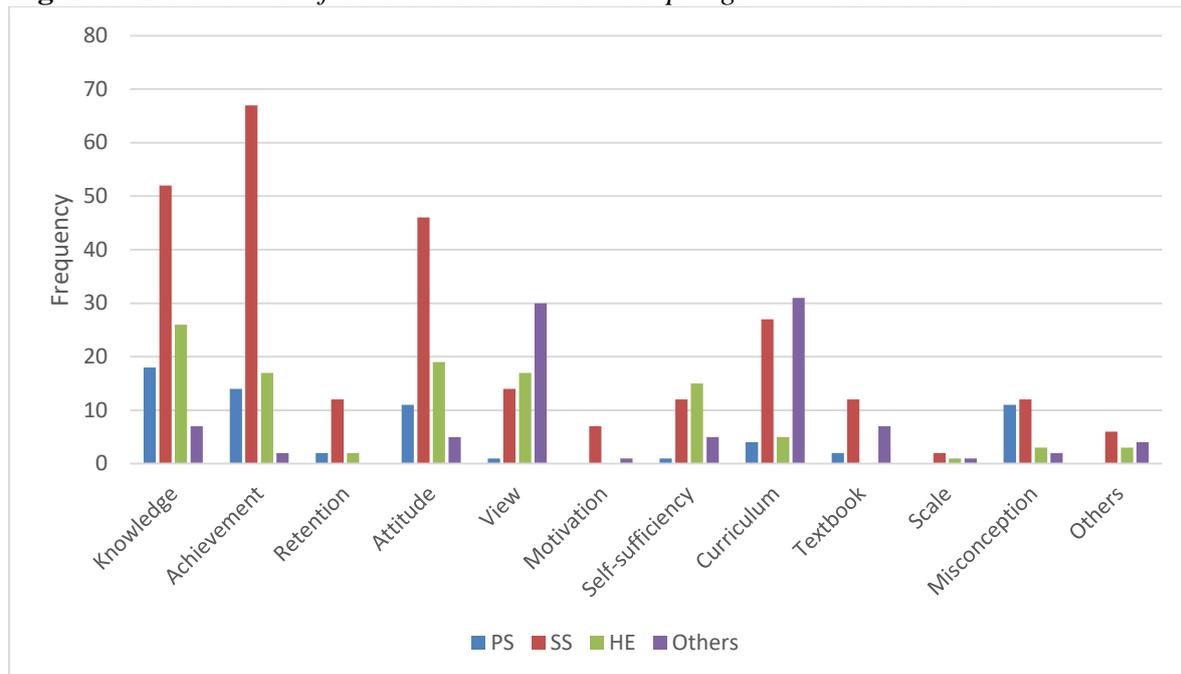
“CEc”; “Hacettepe” in “Ev”; “Gazi” in “PTPC” and “HH”; “Gazi”, “Hacettepe” and “Necmettin Erbakan” in “STinPs”; “Atatürk” and “Gazi” in “NS” and “MSSs”; “Atatürk”, “Balıkesir” and “Gazi” in “DS”, “RS” and “US”; “Balıkesir” in “CS”; “Atatürk” and “Marmara” in “EV”; and to “Gazi” in “Ph” and “CR” topics (Table 4).

Table 4. Distribution of the topics on university basis

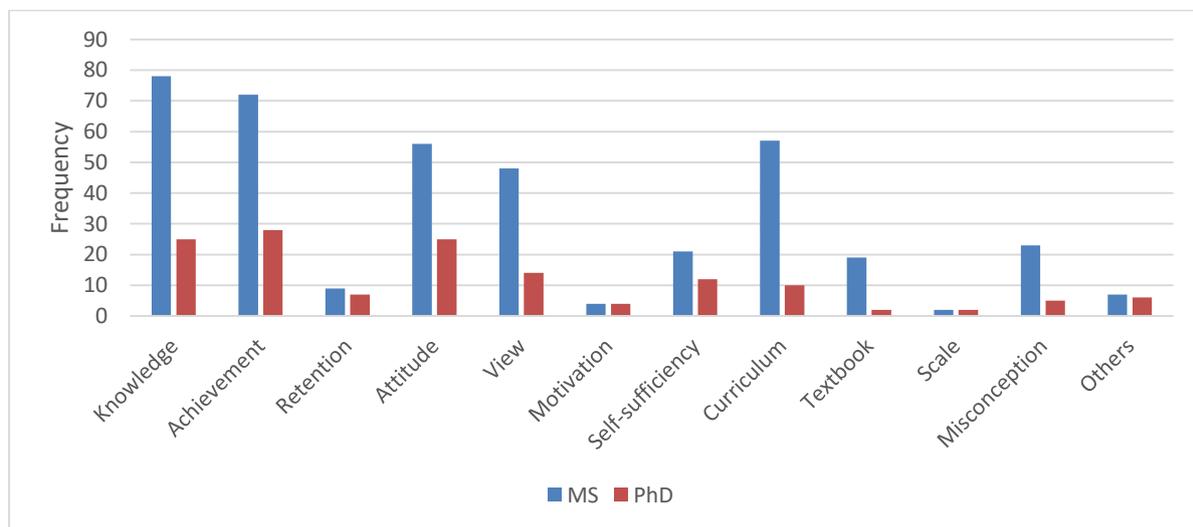
	Atatürk	Balıkesir	Dicle	Gazi	Hacettepe	KTU	Marmara	Necmettin Erbakan	Yüzüncü Yıl	Dokuz Eylül
NSKB	0	23.1(3)**	0	38.5(5)*	0	0	0	15.4(2)	7.7(1)	15.4(2)
CFLT	0	0	0	0	0	0	0	0	0	100(1)*
CLT	12.5(2)	18.8(3)**	6.2(1)	37.5(6)*	18.8(3)**	6.2(1)	0	0	0	0
ST	16.7(1)**	0	0	50(3)*	0	0	0	16.7(1)**	0	16.7(1)**
DCLT	17.6(3)**	11.8(2)	5.9(1)	35.3(6)*	0	11.8(2)	5.9(1)	0	5.9(1)	5.9(1)
KF	11.8(2)	5.9(1)	0	35.6(6)*	23.5(4)**	5.9(1)	0	0	5.9(1)	11.8(2)
CD	27.3(3)*	9.1(1)**	9.1(1)**	27.3(3)*	0	27.3(3)*	0	0	0	0
Ce	18.2(2)**	9.1(1)	9.1(1)	45.5(5)*	0	0	9.1(1)	0	9.1(1)	0
RGH	0	20(1)**	20(1)**	0	0	40(2)*	0	0	0	20(1)**
MGB	0	8.3(2)	0	70.8(17)*	12.5(3)**	0	4.2(1)	0	0	4.2(1)
FGP	16.7(1)**	16.7(1)**	0	33.3(2)*	33.3(2)*	0	0	0	0	0
BPH	13.3(2)**	13.3(2)**	0	20(3)*	13.3(2)**	13.3(2)**	20(3)*	0	0	6.7(1)
CEI	4.3(2)	6.4(3)	0	42.6(20)*	14.9(7)**	14.9(7)**	6.4(3)	6.4(3)	0	4.3(2)
Eec	4.2(1)	4.2(1)	0	41.7(10)*	8.3(2)	8.3(2)	12.5(3)	0	0	20.8(5)**
Bs	25(1)**	0	0	50(2)*	0	25(1)**	0	0	0	0
Cec	25(1)**	0	0	50(2)*	0	25(1)**	0	0	0	0
PE	25(1)**	0	0	50(2)*	0	25(1)**	0	0	0	0
Be	0	0	0	0	0	0	0	0	0	0
OL	0	0	0	0	0	0	0	0	0	0
Ev	20(1)**	0	0	0	40(2)*	0	20(1)**	20(1)**	0	0
PTPC	25(1)**	0	0	50(2)*	0	0	0	0	25(1)**	0
HH	25(1)**	0	0	50(2)*	0	25(1)**	0	0	0	0
STinPs	0	0	0	33.3(1)*	33.3(1)*	0	0	33.3(1)*	0	0
AT	0	0	0	0	0	0	0	0	0	0
NS	50(4)*	0	0	50(4)*	0	0	0	0	0	0
DS	28.6(2)*	28.6(2)*	0	28.6(2)*	0	0	14.3(1)**	0	0	0
RS	33.3(2)*	33.3(2)*	0	33.3(2)*	0	0	0	0	0	0
US	33.3(2)*	33.3(2)*	0	33.3(2)*	0	0	0	0	0	0
CS	22.2(2)**	33.3(3)*	0	22.2(2)**	0	11.1(1)	11.1(1)	0	0	0
MSSs	50(2)*	0	0	50(2)*	0	0	0	0	0	0
EV	40(2)*	0	0	0	0	20(1)**	40(2)*	0	0	0
Ph	14.3(1)**	0	0	71.4(5)*	0	0	0	0	0	14.3(1)**
CR	0	0	0	100(1)*	0	0	0	0	0	0
Ch	0	0	0	0	0	0	0	0	0	0

*The highest value with regards to the categories. ** The second highest value. Frequency (f) was in parentheses. Percent (%) was outside of parentheses.

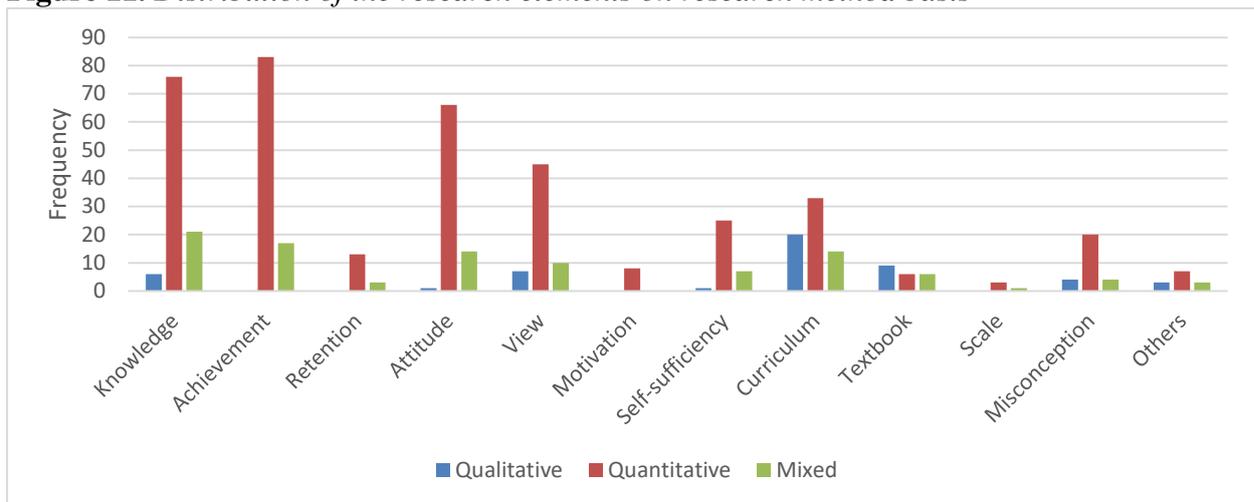
As seen in Figure 9, with regards to the topics studied in different school levels; “Knowledge”, “Achievement” and “Attitude” are studied the most at “SS” level; “Knowledge”, “Achievement”, “Attitude” and “Misconception” are studied the most at “PS” level; “Knowledge”, “Achievement”, “Attitude”, “View” and “Self-sufficiency” are studied the most at “HE” level; “View” and “Curriculum” are studied the most at “Others”. “Retention”, “Motivation”, “Scale” and “Others” are among the elements studied the least.

Figure 9. *Distribution of the research elements on postgraduate theses basis*

As observed in Figure 10; “Knowledge”, “Achievement”, “Attitude” and “Curriculum” are the elements investigated the most in master’s degree and doctoral studies. In “Retention”, “Motivation”, “Scale”, and “Others” categories, the values are similar to each other in master’s degree and doctoral studies. Yet; master’s degree studies are observed to take higher values in other elements.

Figure 10. *Distribution of the research elements on level basis*

As can be seen in Figure 11, the “Quantitative” method is the most preferred method in research elements and the “Qualitative” method is the least preferred one. The “Qualitative” method is not utilized by researchers in “Achievement”, “Retention”, “Motivation” and “Scale” elements; the “Mixed” method is not utilized in “Motivation” and “Scale” elements. On the other hand; in “Textbook” element exclusively, the “Qualitative” method is preferred more than the “Quantitative” method.

Figure 11. Distribution of the research elements on research method basis

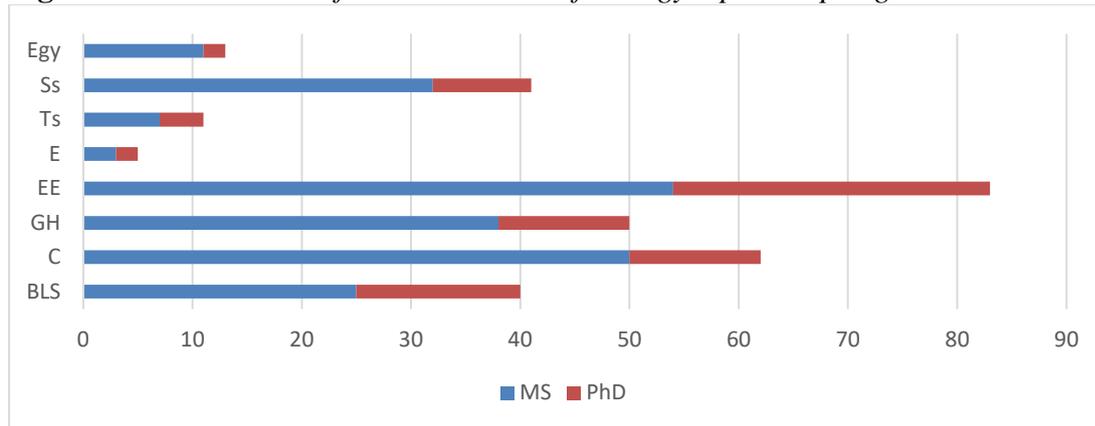
With regards to the type of publication (MS and PhD), the highest values belong to “MS” category in all of the elements. With regards to the levels the researches are intended for (PS, SS, HE and Others); the highest values belong to “SS” level in the “Knowledge”, “Achievement”, “Retention”, “Attitude”, “Motivation”, “Textbook”, “Scale”, “Misconception” and “Others” elements; “HE” level in the “Self-sufficiency” element and to “Others” in the “Curriculum” element (Table 5).

With regards to the research method (Qualitative, Quantitative and Mixed); the “Quantitative” method is observed to take the highest values in all the elements except for the “Textbook” and the “Qualitative” method is observed to take the highest value in the “Textbook” element (Table 5).

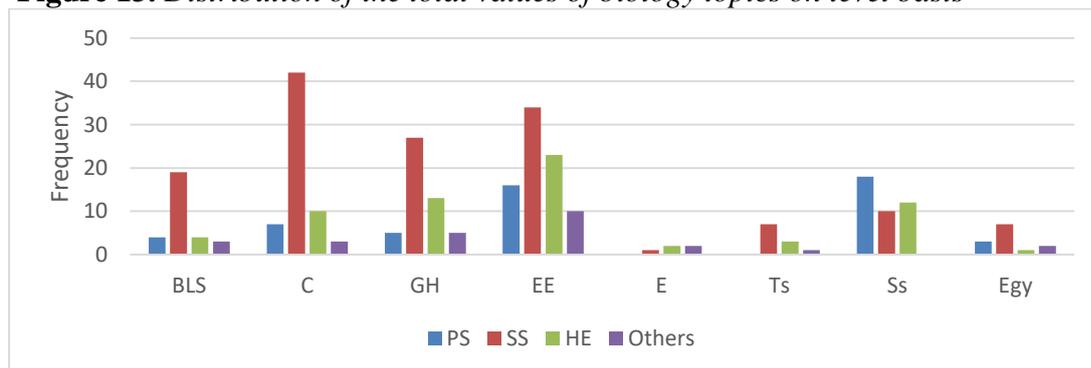
Table 5. General Distribution of the research elements

	MS	PhD	PS	SS	HE	Others	Qualitative	Quantitative	Mixed
Knowledge	75.7(78)	24.3(25)	17.5(18)	50.5(52)*	25.2(26)**	6.8(7)	5.8(6)	73.8(76)*	20.4(21)
Achievement	72(72)	28(28)	14(14)	67(67)*	17(17)**	2(2)	0	83(83)*	17(17)
Retention	19(9)	15(7)	4(2)	25(12)*	4(2)	0	0	27(13)*	6(3)
Attitude	69.1(56)	30.9(25)	13.6(11)	56.8(46)*	23.5(19)**	6.2(5)	1.2(1)	81.5(66)*	17.3(14)**
View	77.4(48)	22.6(14)	1.6(1)	22.6(14)	27.4(17)**	48.4(30)*	11.3(7)	72.6(45)*	16.1(10)**
Motivation	50(4)	50(4)	0	87.5(7)*	0	12.5(1)**	0	100(8)*	0
Self-sufficiency	63.6(21)	36.4(12)	3(1)	36.4(12)**	45.5(15)*	15.2(5)	3(1)	75.8(25)*	21.2(7)**
Curriculum	85.1(57)	14.9(10)	6(4)	40.3(27)**	7.5(5)	46.3(31)*	29.9(20)**	49.3(33)*	20.9(14)
Textbook	90.5(19)	9.5(2)	9.5(2)	57.1(12)*	0	33.3(7)**	42.9(9)*	28.6(6)**	28.6(6)**
Scale	50(2)	50(2)	0	50(2)*	25(1)**	25(1)**	0	75(3)*	25(1)**
Misconception	82.1(23)	17.9(5)	39.3(11)**	42.9(12)*	10.7(3)	7.1(2)	14.3(4)**	71.4(20)*	14.3(4)**
Others	53.8(7)	46.2(6)	0	46.2(6)*	23.1(3)	30.8(4)**	23.1(3)**	53.8(7)*	23.1(3)**

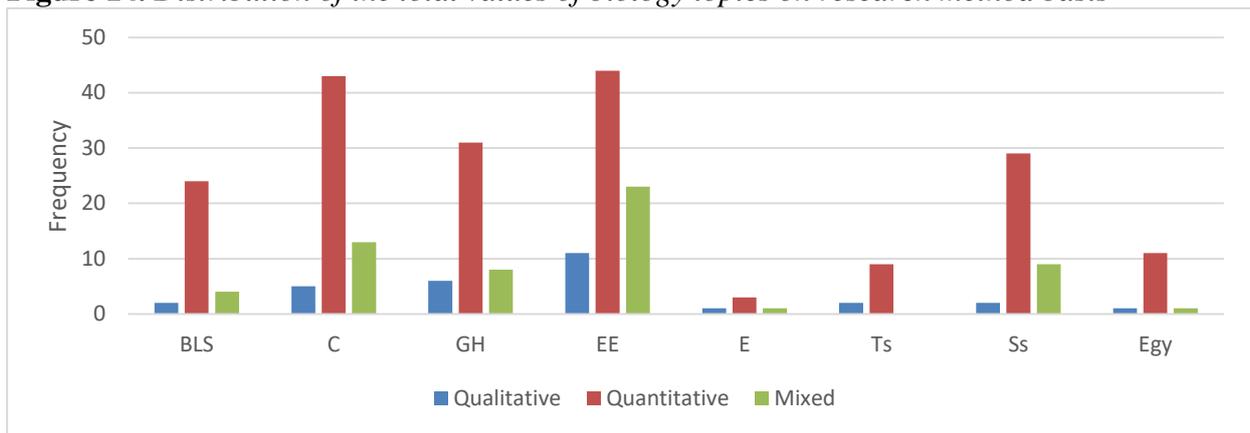
*Frequency (f) was in parentheses. Percent (%) was outside of parentheses.

Figure 12. *Distribution of the total values of biology topics on postgraduate theses basis*

As seen in Figure 12, the most preferred topics with regards to “MS” and “PhD” categories are “BLS” (MS 62%, PhD 38%), “C” (MS 81%, PhD 19%), “GH” (MS 76%, PhD 24%), “EE” (MS 65%, PhD 35%) and “Ss” (MS 78%, PhD 22%). “E” (MS 60%, PhD 40%), “Ts” (MS 64%, PhD 36%) and “Egy” (MS 85%, PhD 15%) are among the least preferred topics. “EE” (MS 65%, PhD 35%) is the most preferred topic both in “MS” and in “PhD” categories. The least preferred topic is “E”. But all of total values of biology topics have more score in MS. In all of categories PhD has lower values.

Figure 13. *Distribution of the total values of biology topics on level basis*

As seen in Figure 13, with regards to the topics that are preferred the most, it can be stated that “EE” is preferred in at all levels (PS 19%, SS 41%, HE 28 and Others 12%). “Ss” and “EE” are among the most preferred topics at “PS” level; “C”(PS 11%, SS 68%, HE 16 and Others 5%), GH” (PS 10%, SS 54%, HE 26 and Others 10%) and “EE” are among the most preferred topics at “SS” level and “EE” is among the most preferred topics in “Others” category. “E”,(PS 0%, SS 20%, HE 40 and Others 40%) “Ts” (PS 19%, SS 41%, HE 28 and Others 12%) and “Egy” (PS 0%, SS 64%, HE 27 and Others 9%) are among the least preferred topics at all levels.

Figure 14. Distribution of the total values of biology topics on research method basis

With regards to the research methods; the “Qualitative”, “Quantitative” and “Mixed” research methods have taken the highest values in “EE” category. The highest values in all topics belong to the “Quantitative” method (BLS 80%, C 71%, GH 69, EE 56%, E 60%, Ts 82%, Ss 72% and Egy %84). The lowest value belongs to “E” for all the three methods. “Ts” is observed to be the category which has not included any of the three methods in the studies. “BLS” (Qualitative 7%, Quantitative 80% and Mixed 13%), “E” (Qualitative 20%, Quantitative 60% and Mixed 20%), “Ts” (Qualitative 18%, Quantitative 82% and Mixed 0%) and “Egy” (Qualitative 8%, Quantitative 84% and Mixed 8%) are observed to be among the topics in which the “Qualitative” and “Mixed” research methods have taken the lowest values.

With regards to the type of publication (MS and PhD) and the research methods (Qualitative, Quantitative and Mixed); “MS” and THE “Quantitative” method have taken the highest values in all topics. With regards to the levels (PS, SS, HE and Others); the highest values belong to; “SS” in “NSKB” and “CLT”; “PS” in “CFLT”; “SS” and “HE” in “ST” and “KF”; “SS” in “DCLT”, “CD”, “Ce”, “RGH”, “MGB”, “FGP”, “BPH”, “CEI” and “EEc”; “PS” and “HE” in “BS”, “CEc” and “Pe”; “SS” and “HE” in “Ev”; “SS” in “PTPC” and “HH”; “HE” in “STinPs”; “SS” and “HE” in “CS”; “SS” in “EV” and “Ph” and to “PS” in “CR”.

Table 6. Distribution of the biology topics on the basis of all variables

	MS	PhD	PS	SS	HE	Others	Qualitative	Quantitative	Mixed
NSKB	76.9(10)	23.1(3)	0	69.2(9)*	23.1(3)**	7.7(1)	7.7(1)	69.2(9)	23.1(3)
CFLT	100(1)	0	100(1)*	0	0	0	0	100(1)	0
CLT	87.5(14)	12.5(2)	18.8(3)**	62.5(10)*	6.2(1)	12.5(2)	6.2(1)	87.5(14)	6.2(1)
ST	83.3(5)	16.7(1)	0	83.3(5)*	16.7(1)*	0	0	66.7(4)	33.3(2)
DCLT	70.6(12)	29.4(5)	11.8(2)**	82.4(14)*	0	5.9(1)	5.9(1)	76.5(13)	17.6(3)
KF	82.4(14)	17.6(3)	11.8(2)**	41.2(7)*	41.2(7)*	5.9(1)	17.6(3)	64.7(11)	17.6(3)
CD	81.8(9)	18.2(2)	27.3(3)**	63.6(7)*	9.1(1)	0	0	72.7(8)	27.3(3)
Ce	90.9(10)	9.1(1)	0	81.8(9)*	9.1(1)**	9.1(1)**	9.1(1)	72.7(8)	18.2(2)
RGH	80(4)	20(1)	40(2)**	60(3)*	0	0	0	80(4)	20(1)
MGB	79.2(19)	20.8(5)	0	50(12)*	37.5(9)**	12.5(3)	16.7(4)	66.7(16)	16.7(4)
FGP	50(3)	50(3)	0	50(3)*	33.3(2)**	16.7(1)	16.7(1)	50(3)	33.3(2)
BPH	80(12)	20(3)	20(3)**	60(9)*	13.3(2)	6.7(1)	6.7(1)	73.3(11)	20(3)
CEI	70.2(33)	29.8(14)	14.9(7)	46.8(22)*	21.3(10)**	17(8)	10.6(5)	68.1(32)	21.3(10)
Eec	62.5(15)	37.5(9)	12.5(3)	50(12)*	29.2(7)**	8.3(2)	12.5(3)	58.3(14)	29.2(7)
Bs	50(2)	50(2)	50(2)*	0	50(2)*	0	25(1)	25(1)	50(2)
Cec	50(2)	50(2)	50(2)*	0	50(2)*	0	25(1)	25(1)	50(2)
PE	50(2)	50(2)	50(2)*	0	50(2)*	0	25(1)	25(1)	50(2)

Be	0	0	0	0	0	0	0	0	0
OL	0	0	0	0	0	0	0	0	0
Ev	60(3)	40(2)	0	20(1)**	40(2)*	40(2)*	20(1)	60(3)	20(1)
PTPC	50(2)	50(2)	0	75(3)*	0	25(1)**	25(1)	75(3)	0
HH	50(2)	50(2)	0	75(3)*	25(1)**	0	0	100(4)	0
STinPs	100(3)	0	0	33.3(1)**	66.7(2)*	0	33.3(1)	66.7(2)	0
AT	0	0	0	0	0	0	0	0	0
NS	62.5(5)	37.5(3)	25(2)**	37.5(3)*	37.5(3)*	0	0	75(6)	25(2)
DS	71.4(5)	28.6(2)	71.4(5)*	14.3(1)**	14.3(1)**	0	14.3(1)	71.4(5)	14.3(1)
RS	100(6)	0	50(3)*	16.7(1)	33.3(2)*	0	0	83.3(5)	16.7(1)
US	66.7(4)	33.3(2)	33.3(2)**	16.7(1)	50(3)*	0	0	83.3(5)	16.7(1)
CS	88.9(9)	11.1(1)	55.6(5)**	22.2(2)*	22.2(2)*	0	11.1(1)	66.7(6)	22.2(2)
MSSs	75(3)	25(1)	25(1)**	50(2)*	25(1)**	0	0	50(2)	50(2)
EV	80(4)	20(1)	0	80(4)*	20(1)**	0	0	80(4)	20(1)
Ph	85.7(6)	14.3(1)	28.6(2)**	42.9(3)*	0	28.6(2)**	14.3(1)	85.7(6)	0
CR	100(1)	0	100(1)*	0	0	0	0	100(1)	0
Ch	0	0	0	0	0	0	0	0	0

*Frequency (f) was in parentheses. Percent (%) was outside of parentheses.

CONCLUSION AND DISCUSSION

This study was conducted to reveal out which biology topics were preferred the most in Turkey in postgraduate theses; and to make comparisons of the university teaching biology, the types of publications, the research methods and research elements. It was also aimed to find out the intensity of biology topics dealt with in primary school and secondary school curricula with regards to course hours on the basis of both the grades and the biology topics (MEB 2013a and MEB 2013b) to check whether there was a congruity with the postgraduate theses or not (Figure 15 and Figure 16).

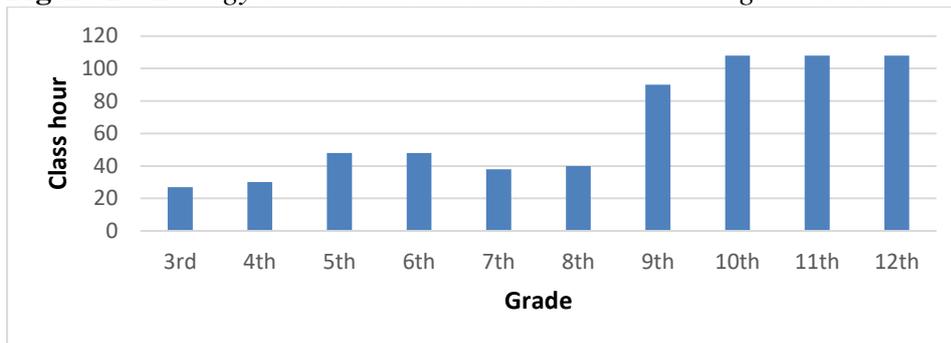
“Except for the “Gazi” university, all the universities were observed to make few theses on biology topics between 1989 and 2015 (Figure 1). There are two reasons for that; first of all, educational researches have gained speed in our country after 2000 and reforms have been made in these areas, as well. The second reason is that, in some of the universities teaching biology, the postgraduate studies were left to the preference of the students. The students can study biology or biology education at postgraduate level. On the other hand, some universities have different approaches. For example, in biology education at education faculty, postgraduate studies are made obligatory. The ones intending to study the area are directed to the institutes of natural sciences.

It was observed that; before 2000, MS theses were mostly done and PhD theses were fewer (Figure 2). “SS” level was mostly preferred as for the research group (Figure 3) and the “Quantitative” method was mostly preferred as for the research method (Figure 4). The “Quantitative” method had been most frequently utilized in several studies that had been conducted in the past (Chang and Hsieh, 1997; White, 1997; Tsai and Wen, 2005; Çalık et. al., 2008; Erdoğan, Marcinkowski and Ok, 2009; Şimşek et. al., 2009; Lee, Wu and Tsai, 2009; Çiltaş, Güler and Sözbilir, 2012; Doğru et. al., 2012; Gül and Sözbilir, 2015). The quantitative methods are preferred more in researches as they are easier to express statistically and to compare with past data (Çalık et. al., 2008). On the other hand; there has been increase in all variables after the year 2000. Particularly the researches on natural sciences are new in Turkey; this can be considered as the reason why the amount of publications was fewer before the year 2000. When the tendency towards researches on natural sciences increased after the year 2000, there has been increase in the amount of studies conducted by researchers. This tendency has brought about variability in the research methods utilized. As researches on

natural sciences have found place in Turkey since 1990ies, they have improved in terms of quality and increased in terms of quantity. These improvements have gained speed particularly after the year 2000 (Çalık et. al., 2008; Gül and Sözbilir, 2015). On the other hand; with the reforms made in education faculties after 1990ies, academicians in these faculties tended towards educational researches (Çiltaş, Güler and Sözbilir, 2012). This was particularly reflected on the university entrance exam. With the improvements made in education after 2000 and the increase in the amount of researches, there has been a change in the quality of the questions asked in university entrance exams. While the questions before that date were mainly focused on certain topics and they were commonly based on knowledge, since 2000 the quality of questions has been improved and the questions have been distributed homogenously in terms of topics (Sülün, 2002). “SS” level is one of the research groups mostly preferred by the researchers (Figure 3). This can be due to the fact that; the number of biology course hours is greater in secondary schools with a greatest share among all the classes (Figure 15). In a similar study, biology topics were reported to be more comprehensive and in greater amount in secondary schools (Gül and Sözbilir, 2015).

The number of biology course hours is observed to be greater in secondary schools. While the number of course hours is 80 and above in secondary school, the number is below 60 in primary schools (Figure 15).

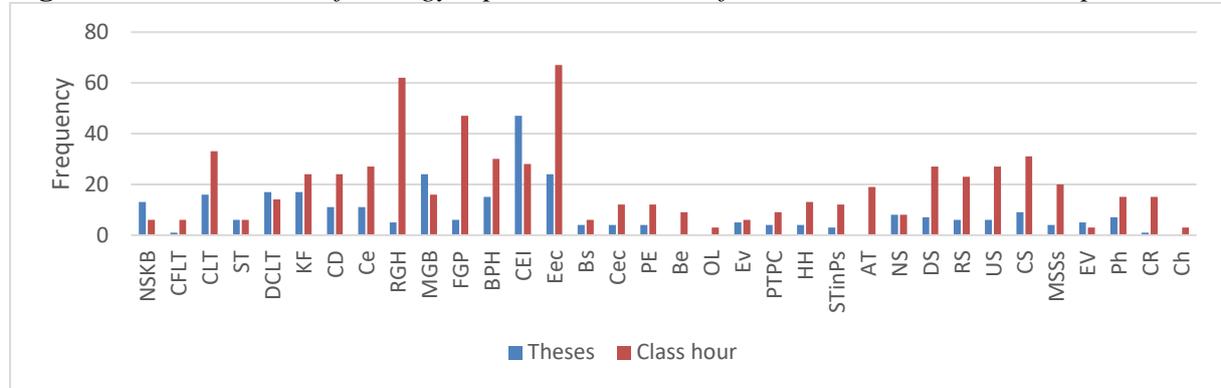
Figure 15. Biology class hours and their distribution on grade basis



It was revealed out that; the environment, cell and variety of the living beings were among the topics that were studied the most in researches while the tissues, evolution, systems and power transformations were studied the least. This can be mainly related to the intensity of the topics taught in primary and secondary schools. There is a similarity between the biology topics given in secondary schools and the topics preferred the most by the researchers (Figure 7, Table 4, Figure 12, Figure 13, Figure 14 or Table 6). Considering the general distribution of the biology topics taught in curricula (Figure 16); the environment, cell, variety of the living beings and classification of the living beings are among the topics taught in greater number of course hours both in primary and in secondary schools. The evolution, tissues, systems and power transformations topics are the ones with fewer course hours. The researches made on less favourable topics such as the reproduction, genetics, tissues, system and power transformations were revealed out to be the topics the students were having the most difficulty in understanding. Insufficiency of the materials, inability to make associations with the daily life and the fact that the topics are based on memorization were reported to be the main reasons for this (Bahar, Johnstone and Hansell, 1999; Özatlı and Bahar, 2010; Kurt and Temelli, 2011; Güneş and Güneş, 2005). To relieve the students from memorization and to enable meaningful learning is one of the most significant objectives of biology education (Özatlı and Bahar, 2010). On the other hand; unless the biology topics making up the most of the biology topics and in which students are having difficulty are included in researches and high quality materials are put into the use of teachers, meaningful

learning cannot be achieved. According to Hong, Shim and Chang (1998); the interest of the individuals on the subject has a great impact on their success. Most probably, students will have a negative attitude against the topics they are having difficulty in understanding. This would bring about inattention and failure.

Figure 16. Distribution of biology topics on the basis of course hours and theses topics



“Knowledge”, “Achievement” and “Attitude” are among the elements that are studied the most, “Retention”, “Motivation”, “Scale” and “Textbook” are among the elements that are studied the least by the researchers (Table 5 and Table 3). With the increase observed in the intensity of science education in recent years, academic success and attitudes of the students have started to be included among the elements that are studied the most (Tsai and Wen 2005). The purpose of education can be stated to be one of the most significant reasons that “knowledge”, “success” and “attitude” are studied the most. This is because; students are expected to develop favourable behaviours through education and to learn information related to the subjects they are studying in the classroom environment (Uitto, Juuti, Lavonen and Meisalo, 2006). In other words, it is significant for the teachers, researchers and educational institutions to find out whether the individuals have developed the expected behaviors and to evaluate these.

In conclusion; the amount and variety of the researches on biology education have increased after the year 2000. On the other hand, certain biology topics have been preferred whereas the topics the students find boring, have difficulty in understanding and the teachers have difficulty in explaining (the systems, tissues, power transformations) are generally neglected in the researches. The topic students have difficulty in learning have small share in the curriculum compared to the other topics.

REFERENCES

- Akaydın, G., Güler, H. M., & Mülayim, H. (2000). The current situation of our science labs in terms of equipment in our lycees [Liselerimizin biyoloji laboratuvar araç ve gereçleri bakımından durumu]. *Hacettepe University Journal of Education*, 19, 1-4.
- Altunoğlu, D. B., & Atav, E. (2005). Teacher expectations for a more efficient biology instruction [Daha etkili bir biyoloji öğretimi için öğretmen beklentileri]. *Hacettepe University Journal of Education*, 28, 19-28.
- Bahar, M., Johnstone, H. A., & Hansell, H. M. (1999). Revisiting learning difficulties in biology. *Journal of Biological Education*, 33 (2), 84-86, DOI: 10.1080/00219266.1999.9655648.

- Chang, H.-Y., Chang, Y.-C., & Tseng, H.-Y. (2010). Trends of Science Education Research: An Automatic Content Analysis. *J.Sci. Educ. Technol.*, 19, 315–331, DOI 10.1007/s10956-009-9202-2.
- Chang, L.-P., & Hsieh, N.-P. (1997). A qualitative review of doctoral dissertations on management in Taiwan. *Higher Education*, 33, 115–136.
- Çalık, M., Ünal, S., Coştu, B., & Karataş, Ö. F. (2008). Trends in Turkish Science Education. *Essays in Education*, 23-45, (Special Edition).
- Çiltaş, A., Güler, G., & Sözbilir, M. (2012). Mathematics Education Research in Turkey: A Content Analysis Study. *Educational Sciences: Theory & Practice*, 12(1), 565-580.
- Delibaş, H., & Babadoğan, C. (2009). A Comparison of Biology Teacher Education Programs in Germany, England and Turkey. *Elementary Education Online*, 8(2), 556-566.
- Doğru, M., Gençosman, T., Ataalkın N. A., & Şeker, F. (2012). Fen Bilimleri Eğitiminde Çalışılan Yüksek Lisans ve Doktora Tezlerinin Analizi. *Journal of Turkish Science Education*, 9 (1), 49-64.
- Elo, S., & Kyngas, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62 (1), 107-115, doi: 10.1111/j.1365-2648.2007.04569.x
- Erdem, D. (2011). Türkiye’de 2005–2006 Yılları Arasında Yayımlanan Eğitim Bilimleri Dergilerindeki Makalelerin Bazı Özellikler Açısından İncelenmesi: Betimsel Bir Analiz. *Journal of Measurement and Evaluation in Education and Psychology*, 2(1), 140-147.
- Erdoğan, M., Marcinkowski, T., & Ok, A. (2009). Content analysis of selected features of K-8 environmental education research studies in Turkey, 1997–2007. *Environmental Education Research*, 15 (5), 525-548, DOI: 10.1080/13504620903085776.
- Gül, Ş., & Sözbilir, M. (2015). Biology education research trends in Turkey. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(1), 93-109.
- Güneş, H. M., & Güneş, T. (2005). Difficulties and their Reasons In Learning Biology Concepts In Primary School Students. *Gazi Üniversitesi Kırşehir Eğitim Fakültesi*, 6 (2), 169-175.
- Harurluoğlu, Y., & Kaya, E. (2009). Biyoloji Öğretmen Adaylarının Biyoloji Öğretimine Yönelik Öz-Yeterlik İnançları. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, XXII (2), 481-496.
- Hong, L.-J., Shim, C.-K., & Chang, K.-N. (1998). A study of Korean middle school students’ interests in biology and their implications for biology education. *International Journal of Science Education*, 20 (8), 989-999, DOI: 10.1080/0950069980200806.
- Kahyaoğlu, H., & Yavuzer, Y. (2004). The Level of knowledge of pre-service teachers related to units used in science lessons given to the 5 th elementary school [Öğretmen Adaylarının ilköğretim 5. Sınıf Fen Bilgisi Dersindeki Ünitelere ilişkin Bilgi Düzeyleri]. *Elementary Education Online*, 3(2), 26-34.
- Kaya, E., & Gürbüz, H. (2002). The Views Of The High Schools and Vocational High Schools Students On The Problems Of Biology Teaching [Lise Ve Meslek Lisesi Öğrencilerinin Biyoloji Öğretiminin Sorunlarına İlişkin Görüşleri]. *Erzincan Eğitim Fakültesi Dergisi*, 4 (2), 11-21.
- Köseoğlu, P., & Soran, H. (2005). Teachers’ Competency Regarding The Use Of Instructional Tools In Teaching Biology [Biyoloji Dersinde Araç-Gereç Kullanımı Açısından Öğretmen Yeterlilikleri]. *Hacettepe University Journal of Education*, 28, 150-158.
- Krell, M., & Krüger, D. (2015). Testing Models: A Key Aspect to Promote Teaching Activities Related to Models and Modelling in Biology Lessons?. *Journal of Biological Education*, 1-14. DOI: 10.1080/00219266.2015.1028570.

- Kurt, M., & Temelli, A. (2011). Effect of Multiple Intelligence Theory Applied for Reproduction Systems on Student Academic Success. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi*, 12 (1), 73-84.
- Lee, H.-M., Wu, T.-Y., & Tsai, C.-C. (2009). Research Trends in Science Education from 2003 to 2007: A content analysis of publications in selected journals, *International Journal of Science Education*, 31(15), 1999-2020, DOI:10.1080/09500690802314876.
- Lin, C.-T., Lin, J.-T., & Tsai, C.-C. (2014). Research Trends in Science Education from 2008 to 2012: A systematic content analysis of publications in selected journals. *International Journal of Science Education*, 36 (8), 1346-1372, DOI:10.1080/09500693.2013.864428
- MEB (2013a). Primary School Science (3rd, 4th, 5th, 6th, 7th and 8th grades) curriculum [İlköğretim Kurumları (İlkokullar Ve Ortaokullar) Fen Bilimleri Dersi (3, 4, 5, 6, 7 ve 8. Sınıflar) Öğretim Programı], <http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72>, Retrieved on 13 November 2015.
- MEB (2013b). Secondary School Biology (9th, 10th, 11th and 12th grades) Curriculum [Ortaöğretim Biyoloji Dersi (9, 10, 11 ve 12. Sınıflar) Öğretim Programı], <http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72>, Retrieved on 13 November 2015.
- McMillan, H. J., & Schumacher, S. (2010). *Research in education: Evidence-based inquiry* (7th ed.). Boston: Pearson.
- Ormanci, U., Cepni, S., Deveci, I., & Aydin, O. (2015). A Thematic Review of Interactive Whiteboard Use in Science Education: Rationales, Purposes, Methods and General Knowledge. *Journal of Science Education and Technology*, 24(5), 532-548, doi:10.1007/s10956-014-9543-3.
- Prokop, P., Prokop, M., & Tunnicliffe, D. S. (2007). Is biology boring? Student attitudes toward biology. *Journal of Biological Education*, 42 (1), 36-39. DOI: 10.1080/00219266.2007.9656105.
- Prokop, P., Tuncer, G., & Chudá, J. (2007). Slovakian Students' Attitudes toward Biology. *Eurasia Journal of Mathematics, Science & Technology Education*, 3(4), 287-295.
- Reiss, J. M. (2006). Teacher education and the new biology. *Teaching Education*, 17(2), 121-131. DOI: 10.1080/10476210600680325.
- Roberts, R. (2001). Procedural understanding in biology: the 'thinking behind the doing', *Journal of Biological Education*, 35 (3), 113-117, DOI: 10.1080/00219266.2001.9655758.
- Saka, A., & Akdeniz, R. A. (2006). The development of computer based material about genetic and application according to 5e model [Genetik Konusunda Bilgisayar Destekli Materyal Geliştirilmesi Ve 5e Modeline Göre Uygulanması]. *The Turkish Online Journal of Educational Technology*, 5 (1), 129-141.
- Sülün, A. (2002). Correspondency of biology questions in university entrance exams with biology teacher education programs and possible reasons of failure [Üniversite Seçme Sınavlarında Yöneltilen Biyoloji sorularının ortaöğretim biyoloji öğretim programlarıyla uygunluğu ve bu alandaki başarısızlığın sebepleri], Ph.D. Thesis, Graduate School of Natural and Applied Sciences, Ataturk University.
- Şimşek, A., Özdamar, N., Uysal, Ö., Kobak, K., Berk, C., Kılıçer, T., & Çiğdem, H. (2009). Current Trends in Educational Technology Research in Turkey in the New Millennium. *Educational Sciences: Theory & Practice*, 9 (2), 941-966.
- Tsai, C.-C., & Wen, L. M. (2005). Research and trends in science education from 1998 to 2002: a content analysis of publication in selected journals. *International Journal of Science Education*, 27 (1), 3-14, DOI: 10.1080/0950069042000243727.

- Önder, N., Oktay, Ö., Eraslan, F., Gülçiçek, Ç., Göksu, V., Kanlı, U., Eryılmaz, A., & Güneş, B. (2013). Content Analysis of Physics Education Studies Published in Turkish Science Education Journal from 2004 to 2011. *Journal of Turkish Science Education*, 10 (4), 151-163.
- Özatlı, S. N., & Bahar, M. (2010). Revealing Students' Cognitive Structures Regarding Excretory System By New Techniques. *Abant İzzet Baysal Üniversitesi Dergisi*, 10 (2), 9-26.
- Usak, M., Prokop, P., Ozden, M., Ozel, M., Bilen, K., & Erdoğan, M. (2009). Turkish university students' attitudes toward biology: the effects of gender and enrolment in biology classes. *Journal of Baltic Science Education*, 8 (2), 88-96.
- Uitto, A., Juuti, K., Lavonen, J., & Meisalo, V. (2006). Students' interest in biology and their out-of school experiences. *Journal of Biological Education*, 40 (3), 124-129, DOI: 10.1080/00219266.2006.9656029.
- Yılmaz, M., & Çimen, O. (2008). Self-Efficiency Belief Levels of Biology Education of Master Students Without Thesis In Biology Education [Biyoloji Eğitimi Tezsiz Yüksek Lisans Öğrencilerinin Biyoloji Öğretimi Öz-Yeterlik İnanç Düzeyleri]. *Yüzüncü Yıl Üniversitesi, Eğitim Fakültesi Dergisi*, V (I), 20-29.
- White, R. (1997). Trends in Research in Science Education. *Research in Science Education*, 27(2), 215-221.
- Zhai, J., Jocz, A. J., & Tan, L.-A. (2014). 'Am I Like a Scientist?': Primary children's images of doing science in school. *International Journal of Science Education*, 36 (4), 553-576, DOI: 10.1080/09500693.2013.791958.

Appendix-1

Content Analysis Form (CAF)

1- University:

2- Publication year:

3- Publication type: MS (...) PhD (...)

4- Sample the study was intended for: Primary school (...) Secondary school (...) High education (...) Others(...)

5-Research method: Qualitative (...) Quantitative (...) Mixed (...)

6- Investigated elements	
61-Knowledge	67-Self-sufficiency
62-Achievement	68-Curriculum
63-Retention	69-Textbook
64-Attitude	610-Scale
65-View	611-Misconception
66-Motivation	612-Others

Biology Subjects and Subjects Matters

7- Biology:Life Sciences (BLS)	8-Cell (C)	9- Genetics and Heredity (GH)	10- Environmental Education (EE)	11-Evolution (E)
71-The Nature of Scientific Knowledge and Biology (NSKB) 72- Common Features of Living Things (CFLT) 73Compounds of Living Things (CLT)	81- Substance Transport (ST) 82- Diversity and Classification of Living Things (DCLT) 83- Kingdom and Features (KF) 84- Cell Division (CD) 85-Cell (Ce)	91- Reproduction, Growth and Development (RGH) 92- Modern Genetics and Biotechnology (MGB) 93- From Gene to Protein (FGP) 94-Basic Principle of Heredity (BPH)	101- Current Environmental Issues (CEI) 102- Ecosystem Ecology (EEc) 103- Biomes (Bs) 104- Community Ecology (CEc) 105- Population Ecology (PE)	111-Behavior (Be) 112- The Origin of Life (OL) 113- Evolution (Ev)
12- Tissues (Ts)	13- Systems (Ss)	14- Energy (Egy)	15-Others (if any ...)	
121- Plant tissues and Plant Construction (PTPC) 122- Hormones and Homeostasis (HH) 123- Substance Transport in Plants (STinPs) 124-Animal Tissues (AT)	131- Nervous System (NS) 132- Digestive System (DS) 133- Respiratory System (RS) 134- Urinary System (US) 135- Circulatory System (CS) 136- Muscular and Skeletal Systems (MSSs)	141- Energy and Vitality (EV) 142- Photosynthesis (Ph) 143- Cellular Respiration (CR) 144- Chemosynthesis (Ch)		

The Development of a Three-tier Chemical Bonding Concept Test

Şenol ŞEN¹ , Ayhan YILMAZ²

¹Dr. Hacettepe University, Faculty of Education, 06800, Beytepe-Ankara, Turkey

²Prof. Dr. Hacettepe University, Faculty of Education, 06800, Beytepe-Ankara, Turkey

Received: 26.08.2016

Revised: 12.12.2016

Accepted: 15.02.2017

The original language of article is English (v.14, n.1, March 2017, pp.110-126, doi: 10.12973/tused.10193a)

ABSTRACT

The aim of this study was to develop a three-tier concept test to determine high school students' conceptual understanding in terms of chemical bonding, and at implementing a reliability and validity study for the test. A total of 175 high school students participated in the research. A three-tier test of 15 questions was employed in the study. Expert opinion was consulted for content and face validity. Furthermore, false positive and false negative values were also calculated for content validity, and on average these values were found to be less than 10%. Statistical analyses were performed for construct validity; Cronbach's Alpha coefficient was calculated as 0.74. To perform item analysis, the top 27% of the students and the bottom 27% of the students were determined. As a result, it was found that item difficulty indices ranged between .47 and .77, and that the discrimination indices were above .30. Following the analyses, both item difficulty indices and discrimination indices were found to be adequate. At the end of the study, Chemical Bonding Concept Test was considered to be reliable and valid.

Keywords: Chemical bonding, chemistry education, reliability, three tier tests, validity.

INTRODUCTION

Students bring into the classroom setting their ideas, concepts and interpretations of the actions happening around them to make their learning easy or difficult; and these ideas, concepts and interpretations differ from scientific thoughts, concepts, definitions or ideas (Chandran, Treagust, & Tobin, 1987; Hammer, 1996; Reynolds & Walberg, 1992; Taber, 2009). Champagne, Gunstone, and Klopfer (1985) and West, Fensham, and Garrard (1985) attribute students' having difficulty in science course to their preconceptions about the happenings around them which do not overlap with scientific concepts (As cited in Yürük, 2005). Such preconceptions or ideas are called misconceptions or alternative concepts in the relevant literature (Driver & Easley, 1978; Nakhleh, 1992). Students' misconceptions are among the main factors negatively influencing their learning process (Muller, Sharma, Eklund, & Reimann, 2007). Therefore, identifying students' misconceptions is important in the learning/teaching process.

Interviews (Boo, 1998; Thompson & Logue, 2006), open-ended questions (Çalık & Ayas, 2005; Şekercioğlu & Kocakulah, 2008; Tsaparlis & Papaphotis, 2002), concept maps (Aykutlu & Şen, 2012; Hazel & Prosser, 1994; Şen & Yılmaz, 2013), analogies (Aykutlu &



Şen, 2012), and multiple-choice tests (Şen & Yılmaz, 2012; Uzuntiryaki & Geban, 2005) are used in the literature to determine misconceptions. Although interviews are useful for determining students' misconceptions, the process of collecting the required data and transcribing and analysing them, is too time consuming (Lin, 2004). The fact that concept maps are seldom used often makes it necessary to teach both teachers and students how to use them. Besides, much more time is needed to use concept maps (Kaya, 2008). For this reason, various studies have made use of multiple-choice tests in determining students' misconceptions (Arslan, Çiğdemoğlu, & Moseley, 2012; Çetin-Dindar & Geban, 2011; Gürcay & Gülbaş, 2015; Kanlı, 2015; Peşman & Eryılmaz, 2010; Sreenivasulu & Subramaniam, 2013; Wutti-prom, Sharma, Johnston, Chitaree, & Soankwan, 2009). Due to the fact that multiple-choice tests are easy to administer, mark objectively and are cost-effective, makes them suitable for statistical analyses and therefore more usually preferred by researchers (Wutti-prom *et al.*, 2009).

Determining misconceptions through conventional multiple-choice tests is not a very popular or preferred method. The reason for this is that it is impossible to determine whether students' answers stem from mistakes, lack of knowledge or from another factor apart from those in conventional multiple-choice tests (Voska & Heikkinen, 2000). Yet, it is possible to work with larger sampling groups by using multiple-choice questions. For this reason, it is recommended that multiple-choice questions are prepared in two or three tiers and used accordingly (Eryılmaz & Sürmeli, 2002). In parallel to this recommendation, for the last ten years, two-tier (Artdej, Ratanaroutai, Coll, & Thongpanchang, 2010; Kanlı, 2015; Şen & Yılmaz, 2012) and three-tier (Çetin-Dindar & Geban, 2011; Gürcay & Gülbaş, 2015; Peşman & Eryılmaz, 2010; Şen, 2015; 2016) tests have been used in the literature rather than one-tier multiple-choice tests (Hırça, Çalık, & Akdeniz, 2008). As in conventional tests, content is tested in the first tier of the two and three-tier tests, and the reason for their answers given in the first tier is requested in the second tier (reason tier) (Tan, Goh, Chia, & Treagust, 2002). However, two-tier tests may not be adequate in finding out whether or not students' misconceptions are due to lack of knowledge. Thus, students can be asked to state how confident they are about their answer by adding a third tier (confidence tier). This third tier is used to distinguish between the students' incorrect answers, due to lack of knowledge and their misconceptions. If a student gives incorrect answers in the first and second tiers, despite being sure about his/her answer, it is asserted that the student has a misconception. But if a student is not sure that his/her incorrect answer is correct, the incorrect answer cannot be regarded as a misconception (Peşman, 2005). Misconception tests developed by researchers are also marked by assigning 1 point to correct answers and zero points to incorrect answers - as in achievement tests (Eryılmaz & Sürmeli, 2002).

A review of the literature demonstrated that three-tier tests were developed and used in chemistry subjects, such as Acids and Bases (Çetin-Dindar & Geban, 2011), States of Matter (Kirbulut, Geban, & Beeth, 2010) and Electrochemistry (Şen, 2015; 2016), so as to determine students' misconceptions about chemistry. It was found that three different tests were developed in relation to chemical bonding (Othman, Treagust, & Chandrasegaran, 2008; Peterson & Treagust, 1989; Treagust, 1988); however, none of these was three-tier. Three-tier tests should be developed in order to be able to control external factors, such as lack of knowledge and chance in the process of determining students' misconceptions about chemical bonding; because research has shown that the subject of chemical bonding is difficult and problematic to students and, as a result, students have a large number of misconceptions (Coll, 2008; Coll & Taylor, 2002; Luxford, & Bretz, 2014; Nicoll, 2001; Peterson & Treagust, 1989; Şen & Yılmaz, 2013; Taber, 2011; Tan, & Treagust, 1999; Temel & Özcan, 2016; Yayon, Mamlok-Naaman, & Fortus, 2012). The fact that chemical bonding is an abstract concept and that it is not associated with learners' real life experiences makes it difficult to

learn. As a consequence, that students have misconceptions about chemical bonding is an expected situation (Tan & Treagust, 1999). Using this as a starting point, the aim of this study is to develop a concept test on chemical bonding so as to determine the misconceptions of high school students in this respect, and to perform validity and reliability analyses. Studies available in the literature have shown that students have various misconceptions about the effects of physical changes on intramolecular bonds (Barker, 1995; Mirzalar Kabapınar & Adik, 2005), the formation and structure of ionic bonds (Luxford & Bretz, 2013; 2014), covalent bonds (Luxford & Bretz, 2013; 2014; Peterson & Treagust, 1989; Ünal, Coştu, & Ayas, 2010), metallic bonds (Acar & Tarhan, 2008), and intermolecular and intramolecular bonds (Othman *et al.*, 2008; Peterson & Treagust, 1989; Tan, & Treagust, 1999; Treagust, 1988). When considering these studies, performed in relation to misconceptions in the literature, the importance of a three-tier multiple-choice test to determine students' misconceptions about chemical bonding becomes manifest. For this reason, it is considered that this study will contribute to the literature concerning chemical bonding. The creation in this study of a valid and reliable test will facilitate its use by teachers. For this reason, the objective of this study is analysed using the following research problem:

- Are Chemical Bonding Concept Test results a valid and reliable measure of students' conceptual understanding of chemical bonding?

METHOD

a) Developing a Chemical Bonding Concept Test

The misconceptions concerning chemical bonding that are available in the literature were realized prior to preparing the chemical bonding concept test (Luxford, & Bretz, 2014; Nicoll, 2001; Peterson & Treagust, 1989; Şen & Yılmaz, 2013; Tan, & Treagust, 1999; Temel, & Özcan, 2016). The test prepared included three distractors and a correct answer. The misconceptions in the literature were used to create these distractors. The test contained three tiers and 15 questions. Expert opinion was obtained for the questions prepared for the test. For this purpose, the test was evaluated by two educators of chemistry education and by an expert chemistry teacher. Following the evaluations, the regulations were made, and the test was given its final shape. Subsequently, the test was administered to a group of 25 high school students, and efforts were made to determine whether or not there were any unclear or faulty questions. The questions were revised in accordance with the feedback from the students. Afterwards, the test was applied to 175 high school students for validity and reliability analyses to give it the final shape.

b) The Chemical Bonding Concept Test

Misconceptions included in the chemical bonding concept test (CBCT) are shown in Table 1.

Table 1. *The Distribution of Common Misconceptions included in the Items of CBCT*

Misconceptions	Items
Physical changes affect intramolecular bonds.	1, 2, 3, 11, 13
Compounds containing hydrogen bonds in the structure are in liquid phase.	2
Electrons are transferred from one atom to another during the formation of metal bonds, they can share electron.	4
An ionic bond is an intermolecular bond.	5
Polarity is determined by lone-pair electrons.	7
In all covalent bonds, bond-pair electrons are at equal distances from bonding atoms.	6
Whether a molecule is polar or not does not depend on the polarity of the bonds in that molecule and on the shape of the molecule.	8, 9
Intermolecular bonds are the forces available only in covalent compounds.	10
Energy is required for chemical bonding.	14
Heat and pressure change bond angles and the number of bonds.	11, 13
London forces and hydrogen bonds are available only between polar molecules.	10, 15
Non-bonding electron pairs do not have effects on geometric structure.	12
London forces are not real bonds.	10
London forces are both intermolecular and intramolecular.	10

A sample question included in the test is presented below. In the first tier of the question, the students were asked to explain the correlations between physical changes and heat. Option A is the correct answer, and the other three options are misconceptions. In the second tier, the students were asked to explain the reason for their answer in the first tier. Here, C is the correct option and the other three options are misconceptions. As is apparent from tier three, the students were asked to state whether or not they were sure about their answers.

1a) In which of the following is a change observed during the transition of pure ice from solid phase into liquid phase?

***A)** The strength of hydrogen bonds between molecules

B) Sigma (σ) bond gravity force in the molecule

C) Attractive forces of the covalent bond in the molecule

D) The number of sigma (σ) bonds in the molecule.

1b) Which of the following best explains your reason for your answer?

A) Hydrogen bonds between molecules are not chemical bonds but are only forces/interactions.

B) Sigma bonds are not real chemical bonds but are only weak forces.

***C)** Molecules that have more kinetic energy while ice is melting begin to separate, and hydrogen bonds are preserved to some extent.

D) The attractive forces of covalent bonds need to decrease in order for molecules having more kinetic energy to separate, while ice melting.

1c) About the answer to the above question:

a. I am sure.

b. I am not sure.

* Correct answer.

c) The Study Group

A total of 175 12th grade high school students were included in the study. The average age of the students was 18.02 (SD=.23). Of the participants, 84 were female and 78 were male. Three of them did not state their age, and thirteen did not state their gender. No mark was given to the papers of the students, who answered the first tier of the test but did not answer the second and third tiers, or those who chose only one option for all three tiers and these were excluded from the study. In addition to this, the data for those students who answered only half of the test, leaving the other half untouched, were also not analysed.

d) Data Collection Process

The students were allowed one class hour to answer the 15 questions in the test. They were encouraged to participate on the basis of volunteering, and were assured that the scores they received from the test would not affect their average academic achievement score.

e) Types of Scores according to Students' Answers

The types of scores for the 15 test questions - which were obtained through analyses using the Microsoft Excel program - are shown in Figure 1. Figure 1 was prepared on the basis of the study by Arslan, Çiğdemoğlu and Moseley (2012).

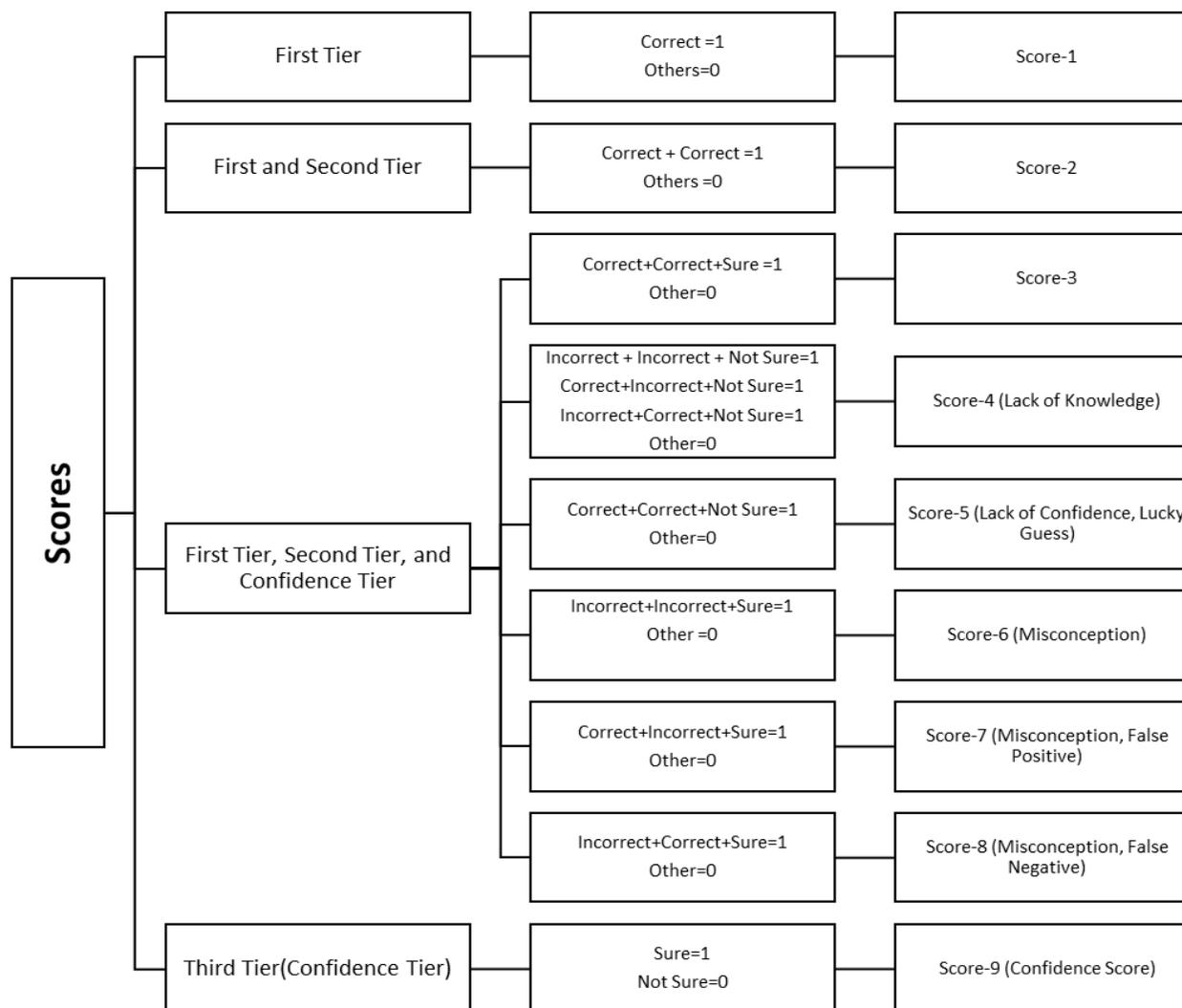


Figure 1. The Coding Procedure for Scores.

Score-1: This type of score was calculated from the students' answers to the first tier. Accordingly, if the students answered correctly, they received 1 point; if they answered incorrectly, they received zero points.

Score-2: This type of score was calculated from the students' answers to the first and second tiers. Accordingly, if they answered correctly in the first and second tiers, they were given 1 point; if they answered incorrectly in both tiers or in one of the tiers, they were given zero points.

Score-3: The students' answers to all three tiers were taken into consideration in this type of score. Accordingly, if the students answered correctly in the first and second tiers and said that they were sure about their answers, they received 1 point. In other cases, they received zero points.

Score-4 (Lack of Knowledge): This score was also calculated by taking the students' answers to all three tiers into consideration. Accordingly, if the students gave incorrect answers in the first and/or second tiers and stated that they were not sure about their answers, they received 1 point. In other cases, they received zero points. The scores the students received for each question demonstrated the students' lack of knowledge, in terms of the relevant concepts. In this way, the classification of each incorrect answer as a misconception was avoided.

Score-5 (Lack of Confidence or Lucky Guess): Here, the students also received 1 point, if they answered correctly in the first and second tiers and if they were not sure about their answers. They received zero points in other cases. In this case, the fact that they had been asked to state whether or not they were sure about their answers, even if they had answered correctly, demonstrated the lucky guess factor.

Score-6 (Misconception): If the students chose the incorrect option as the correct answer, chose the option to explain their reason for the answer and stated in the third tier that they were sure about their answer, then they were considered to have misconceptions. The students received 1 point for each question, as they answered incorrectly in the first and second tiers and they stated in the third tier that they were sure about their answer. In other cases, they received zero points.

Score-7 (Misconception, False Positive): This type of score was calculated when the students answered correctly in the first tier and incorrectly in the second tier, and in the third tier stated that they were sure about their answer. Here, the students received 1 point for each question they answered, and in other cases they received zero points, as the students had given correct answers with incorrect reasoning.

Score-8 (Misconception, False Negative): This type of score was calculated when the students answered incorrectly in the first tier and correctly in the second tier and stated in the third tier that they were sure about their answers. Here, they received 1 point for the question they answered, and they received zero points in other cases. In this case, students gave incorrect answers but with correct reasoning.

Score-9 (Confidence Score): In this type of score, only the students' answers to the third tier were essential. If they marked the option "I am sure," they received 1 point; and if they marked the option "I am not sure," they received zero points.

FINDINGS

The findings obtained in the study are presented under the following headings: results of content, face and construct validity (a), results of item analyses (b), and results of the reliability coefficient (c).

a) Results of Content, Face and Construct Validity

Firstly, the content, face and construct validity of the CBCT was analysed. Initially, expert opinion was obtained in relation to the content and face validity of the test. Two instructors of chemistry education and a chemistry teacher evaluated the test. The experts analysed the questions in the test, in terms of its level of suitability for high school students, its appropriateness to the school curriculum and the psychometric properties in the test, linguistic understandability and instructions for the questions. The items of the CBCT were revised, based on the suggestions and comments submitted by the experts. According to the experts' suggestions and comments, the distractors, linguistic understandability and instructions for the questions were revised. In addition, false positive and false negative values were also calculated for content validity - as is indicated by Hestenes and Halloun (1995). Researchers state that false positive and false negative proportions of less than 10% in a multiple-choice test increase content validity. In this study, the average percentage was found to be 6.51% for false positive and 6.02% for false negative. Therefore, having both false positive and false negative values of less than 10% is the evidence for content validity in this study. Çataloğlu (2002) recommends looking at the correlations between high scores received from the test and confidence levels to check for construct validity, and advocates that students with higher scores should have more self-confidence than those with lower scores. Çataloğlu (2002) and Peşman and Eryılmaz (2010) emphasise that there should be at least a moderate level of positive correlations between the two-tier scores and confidence tier scores. Therefore, the correlations between score-2 and score-9 were analysed in this study. The analyses revealed that there were positive and statistically significant correlations between the two types of scores ($r=.73$, $n=175$, $p<.0005$). Thus, the test was considered to have attained construct validity.

b) Results of Item Analyses

In this study, item analyses were made on the basis of the difference between the group averages of top 27% of the students and the bottom 27% of the students (Wiersma & Jurs, 1990). Item analyses for the items in the test were made on the basis of score-3. Table 2 shows the figures for the results concerning the analyses. Accordingly, item difficulty indices in the first column receive values between .47 and .77. The fact that the item difficulty indices are approximately .50 shows that those items are better and more useful. Crocker and Algina (2006) point out that the item difficulty index is not a property necessary for item quality, but that the value for the item difficulty index should be approximately .50 in order for the total variance and reliability to increase. However, it is stated in the literature that the item difficulty index can also range between .10 and .80 on the condition that it is approximately .50 for many items (Walsh & Betz, 1990). Another study emphasises that the value should be in the .50- .80 interval (Wiersma & Yurs, 1990). Based on this, it was considered that the item difficulty indices for the items in the CBCT were at an adequate level. Another index shown in the table is the discrimination index. It shows to what extent the items discriminate individuals. Ebel (1965) points out that items with a discrimination index of .40 and above are appropriate but that the items with a discrimination index between .30 and .39 can be used in tests with little or no revision (As

cited in, Crocker & Algina, 2006). That the discrimination index for the 15 items in the test was above .30 demonstrated that all the items distinguished between successful and unsuccessful students in terms of chemical bonding. An examination of item point biserial coefficient showed that the value was above .20, except for item 15. Item point biserial coefficients represent the correlations between students' correct answer to an item in the test and their correct answers to the other items in the test. A value above .20 is an acceptable value (Beichner, 1994; Peşman, & Eryılmaz, 2010; Wuttiprom *et al.*, 2009); nevertheless, the minimum acceptable point biserial value for the sample of 175 was found to be .15 according to the formula suggested by Crocker and Algina (2006), depending on the sample size. Thus, item 15 was not removed from the test, as it had the minimum acceptable point biserial value, which was adequate for this study, according to Crocker and Algina (2006). Besides, Crocker and Algina (2006) state that it would be more beneficial to use item discrimination indices in determining problematic items, and that the other values (item difficulty) could also be used in analysing the structure of questions. For this reason, the fact that the item discrimination index for question 15 was greater than .30 was interpreted as the question not being problematic.

Table 2. *Item Analyses*

Item	Item Difficulty Index (Pj)	Item Discrimination index (rj)	Item Variance	Item Standard Deviation	Item point biserial correlation coefficient
M1	.63	.62	.23	.48	.43
M2	.54	.70	.25	.50	.37
M3	.59	.62	.24	.49	.40
M4	.77	.38	.18	.42	.35
M5	.70	.55	.21	.46	.46
M6	.64	.30	.23	.48	.22
M7	.47	.51	.25	.50	.27
M8	.68	.51	.22	.47	.33
M9	.65	.62	.23	.48	.43
M10	.72	.43	.20	.45	.23
M11	.63	.70	.23	.48	.44
M12	.67	.62	.22	.47	.41
M13	.59	.57	.24	.49	.36
M14	.73	.45	.20	.44	.35
M15	.65	.36	.23	.48	.15

c) Results of Reliability Coefficient

Cronbach's Alpha coefficients were calculated separately for score-1, score-2 and score-3. The values calculated for all three types of scores are shown in Table 3. The fact that the Cronbach's alpha coefficients calculated from these three types of scores were above .70 showed that the test was reliable (Cronbach, 1951; George & Mallery, 2003; Pallant, 2005).

Table 3. *Reliability Statistics*

	Cronbach's Alpha	N of Items
Score-3	0,74	15
Score-2	0,77	15
Score-1	0,72	15

Table 4 shows the percentages for false negatives, false positives, lack of knowledge, lack of confidence and misconceptions. These types of scores exhibit the importance of three-tier tests. As is clear from the table, on average the percentage for lack of knowledge is 20.20%. Hence, the value demonstrates that all of the answers in a three-tier test should not be perceived as misconceptions. If one or two-tier tests had been used, students' incorrect answers due to lack of knowledge would have been considered as misconceptions. It is also clear from the table that question two is the item in which students have the most lack of knowledge. The question is about whether or not the electrons in covalent and hydrogen bonds are shared. An examination of lack of confidence percentages shows that the percentage is 16.33%. According to this result, some of the students were not confident about their correct answers or their correct answers were lucky guesses. On examining the values for score-6 (misconceptions), it was found that students had the most frequent misconceptions in the first four questions. It was apparent that the largest number of misconceptions related to question four in particular. This question was related to metallic bonds. It was also found that for question 13 the percentage for false negatives exceeded 10%. Yet, due to the fact that the value was not very high and that the average percentage for false negatives was less than 10%, it was regarded that content validity was attained. Likewise, on examining the false positives for each item, it was found that they were less than 10% except for items 11 and 13. Since false positive values for items 11 and 13 were 10.9% and 10.35, respectively and the average for all items was 6.52%, it was deemed that content validity was attained.

Table 4. Percentages of False Negatives, False Positives, Lack of Knowledge, Lack of Confidence and Misconception

Scores	Item															M	SD
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
Score-4 (Lack of Knowledge)	15	35	22	14	16	11	26	23	20	14	25	16	23	17	26	20,2	6,34
Score-5 (Lack Of Confidence)	42	21	25	9	13	18	47	13	13	7	10	9	9	5	4	16,3	12,8
Score-6 (Misconception)	39	38	35	42	34	33	33	34	30	27	21	31	22	26	21	31,1	6,54
Score -7 (False Positive)	3,4 3	3,4 3	3,4 3	3,4 3	4,5 7	8,5 7	5,1 4	5,1 4	7,4 3	6,2 9	10, 9	8,5 7	10, 3	8,5 7	8,5 7	6,52	2,63
Score -8 (False Negative)	2,2 9	4	4	2,2 9	4,5 7	4,5 7	4	6,2 9	5,7 1	6,8 6	8	8	10, 9	9,1 4	9,7 1	6,02	2,67

Figure 2 shows the percentages for the students' correct answers to the tiers (score-1, score-2, and score-3). Accordingly, it was found that the rate of correct answers to the three tiers of the questions decreased. The reason for this was that the probability of lucky guesses was lower in three-tier tests than in one or two-tier tests. What is important here is the reduction in the number of the students' lucky guesses. According to the graph, while the rate of giving correct answers for the 15 questions in the test was 73% in the first tier, the rate fell to 68% on evaluating the first and second tiers together. In score-3, however, in which the students were required to answer correctly in the first and second tiers and to state that they were sure about their answers, the average percentage fell to 65%.

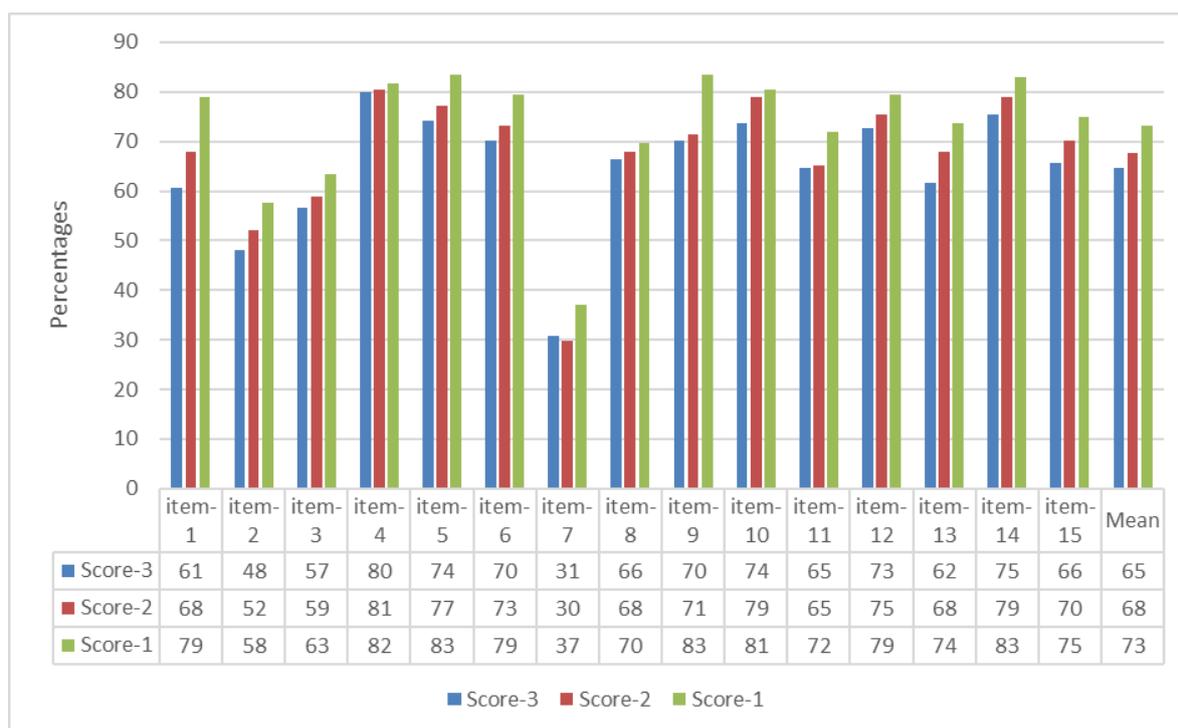


Figure 2. Percentages of Correct Responses in terms of Number of Tiers.

DISCUSSION and CONCLUSION

This study has made an attempt at developing a three-tier concept test so as to determine students' misconceptions about chemical bonding. The test contained 15 questions and three tiers for each question. The first tier of the test was the content tier, and it was directed to finding out what the participating students knew. The second tier was for ascertaining the students' reasons for their answers. The distractors included in these two tiers contained the common misconceptions reported in the literature. In the third tier, on the other hand, the students were asked whether they were sure about their answers in the first two tiers, or not. Prior to the implementation of the test, expert opinion was received for content and face validity. Having received expert opinion, a pilot study was implemented with 25 high school students, and any necessary modifications were made to the test. The false positive and false negative values were calculated for the content validity of the test, and the value was found to be below 10%. Thus, the test was considered to attain content validity. For the construct validity of the test, the correlations between two-tier scores and confidence tier scores were analysed through correlation analysis, and the correlations between the two types of scores were found to be positive and significant. Thus, the test was regarded as having attained construct validity.

In order to carry out item analyses, the top 27% of students and the bottom 27% of students were determined. Consequently, it was observed that the item difficulty indices took on values in the .47-.77 interval. The values demonstrate the relatively difficult and relatively easy questions in the test. The questions in the test, however, were at a medium level of difficulty. That the discrimination indices for all the questions were more than .30 exhibited that all the items in the test discriminated between the successful and unsuccessful students. Cronbach's Alpha coefficients for the test were calculated for score-1, score-2 and score-3. As the values were above .70, this was interpreted as the test being reliable.

At the end of this study, the students' answers to the first three tiers combined (score-3), the answers they gave to the first two tiers combined (score-2), and finally their answers to the first tier only (score-1) were evaluated. In consequence, it was found that students' correct answers to the questions decreased gradually. According to these results, as well as other studies in the literature, three-tier tests are more useful than conventional multiple-choice tests in determining students' misconceptions, their incorrect answers stemming from lack of knowledge and identifying those students who lack confidence, as well as for pointing out their correct answers by means of lucky guesses (Çetin-Dindar & Geban, 2011; Kirbulut *et al.*, 2010; Peşman & Eryılmaz, 2010). The use of such tests in education would also contribute to the recognition of any misconceptions.

There are also restrictions in three-tier tests. The most important of these restrictions is that these tests are not adequate for determining whether students are sure about their answers to the first and second tiers if they state in the third tier that they are not sure. Therefore, four-tier tests can be developed instead of three-tier tests (Caleon & Subramaniam, 2010). In a four-tier test, students can be asked if they are sure about their answers in both tiers.

High school teachers can easily use the CBCT developed in this study because these tests are more reliable and valid than the alternatives. For this reason, they are thought to be effective instruments in determining students' achievement and in uncovering their misconceptions. Furthermore, feedback received by high school teachers will also shed light on the development of such tests.

REFERENCES

- Acar, B., & Tarhan, L. (2008). Effects of cooperative learning on students' understanding of metallic bonding. *Research in Science Education*, 38, 401–420.
- Arslan, H. Ö., Çiğdemoğlu, C., & Moseley, C. (2012). A three-tier diagnostic test to assess preservice teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. *International Journal of Science Education*, 34(11), 1667–1686.
- Artdej, R., Ratanaroutai, T., Coll, R. K., & Thongpanchang, T. (2010). Thai Grade 11 students' alternative conceptions for acid–base chemistry. *Research in Science & Technological Education*, 28(2), 167–183.
- Aykutlu, I., & Şen, A. İ. (2012). Determination of secondary school students' misconceptions about the electric current using a three-tier test, concept maps and analogies. *Education and Science*, 37(166), 275–288.
- Barker, V. (1995). *A longitudinal study of 16-18 year olds' understanding of basic chemical ideas*. Unpublished Doctorate Thesis, Department of Educational Studies, University of York.
- Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American journal of Physics*, 62(8), 750–762.
- Boo, H. K. (1998). Students' understandings of chemical bonds and the energetics of chemical reactions. *Journal of Research in Science Teaching*, 35(5), 569–581.
- Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education*, 32(7), 939–961.
- Çalık, M., & Ayas, A. (2005). A comparison of level of understanding of eighth-grade students and science student teachers related to selected chemistry concepts. *Journal of Research in Science Teaching*, 42(6), 638–667.
- Çataloğlu, E. (2002). *Development and validation of an achievement test in introductory quantum mechanics: The quantum mechanics visualization instrument (QMVI)* (Order No. 3060014). Available from ProQuest Dissertations & Theses Global. (305491729). Retrieved from <http://search.proquest.com/docview/305491729?accountid=11248>.
- Çetin-Dindar, A., & Geban, Ö. (2011). Development of a three-tier test to assess high school students' understanding of acids and bases. *Procedia Social and Behavioral Sciences*, 15, 600–604.
- Chandran, S., Treagust, D. F., & Tobin, K. (1987). The role of cognitive factors in chemistry achievement. *Journal of Research in Science Teaching*, 24(2), 145–160.
- Coll, R. K. (2008). Chemistry learners' preferred mental models for chemical bonding. *Journal of Turkish Science Education*, 5(1), 22–47.
- Coll, R. K., & Taylor, N. (2002). Mental models in chemistry: senior chemistry students 'mental models of chemical bonding. *Chemistry Education: Research and Practice in Europe*, 3(2), 175–184.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt College.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Driver, R., & Easley, J. (1978). Pupils and paradigms: A review of the literature related to concept development in adolescent science students. *Studies in Science Education*, 5(1), 61–84.
- Eryılmaz, A., & Sürmeli, E. (2002). *Üç-aşamalı sorularla öğrencilerin ısı ve sıcaklık konularındaki kavram yanlışlarının ölçülmesi [The assessment of students' misconceptions about heat and temperature concepts by means of three-tier questions]*.

- Paper presented at the 5th National Conference on Science and Mathematics Education. Retrieved June 6, 2016, from <http://users.metu.edu.tr/eryilmaz/TamUcBaglant.pdf>
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference (11.0 update)* (4th ed.). Boston: Allyn & Bacon.
- Gürçay, D., & Gülbaş, E. (2015). Development of three-tier heat, temperature and internal energy diagnostic test. *Research in Science & Technological Education*, 33(2), 197-217.
- Hammer, D. (1996). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. *American Journal of Physics*, 64(10), 1316-1325.
- Hazel, E., & Prosser, M. (1994). First-year university students' understanding of photosynthesis, their study strategies and learning context. *The American Biology Teacher*, 56(5), 274-279.
- Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory. *Physics Teacher*, 33,502-506.
- Hırça, N., Çalık, M., & Akdeniz, F. (2008). Investigating grade 8 students' conceptions of energy and related concepts. *Journal of Turkish Science Education*, 5(1), 75-87.
- Kanlı, U. (2015). Using a two-tier test to analyse students and teachers' alternative concepts in astronomy. *Science Education International*, 26(2), 148-165.
- Kaya, O. N. (2008). A student-centred approach: Assessing the changes in prospective science teachers' conceptual understanding by concept mapping in a general chemistry laboratory. *Research in Science Education*, 38(1), 91-110.
- Kirbulut, D., Geban, O., & Beeth, M. E. (2010, July). Development of a three-tier multiple-choice diagnostic instrument to evaluate students' understanding of states of matter. In *European Conference on Research in Chemical Education (ECRICE)*. Krakow, Poland.
- Lin, S. W. (2004). Development and application of a two-tier diagnostic test for high school students' understanding of flowering plant growth and development. *International Journal of Science and Mathematics Education*, 2(2), 175-199.
- Luxford, C. J., & Bretz, S. L. (2013). Moving beyond definitions: What student-generated models reveal about their understanding of covalent bonding and ionic bonding. *Chemistry Education Research Practice*, 14, 214-222.
- Luxford, C. J., & Bretz, S. L. (2014). Development of the Bonding Representations Inventory to identify student misconceptions about covalent and ionic bonding representations. *Journal of Chemical Education*, 91(3), 312-320.
- Mirzalar Kabapınar, F. (2007). A literature review of student misconceptions on chemical bonding I: Intramolecular bonding. *National Education*, 176, 18-35.
- Muller, D.A., Sharma, M.D., Eklund, J., & Reimann, P. (2007). Conceptual change through vicarious learning in an authentic physics setting. *Instructional Science*, 35(6), 519-533.
- Nakhleh, M. B. (1992). Why some students don't learn chemistry. *Journal of Chemical Education*, 69(3), 191-196.
- Nicoll, G. (2001). A report of undergraduates' bonding misconceptions. *International Journal of Science Education*, 23(7), 707-730.
- Othman, J., Treagust, D. F., & Chandrasegaran, A. L. (2008). An investigation into the relationship between students' conceptions of the particulate nature of matter and their understanding of chemical bonding. *International Journal of Science Education*, 30(11), 1531-1550.
- Pallant, J. (2005). *SPSS Survival Manual: A Step-by-Step Guide to Data Analysis Using SPSS*. Allen and Unwin: Sydney.

- Peşman, H. (2005). *Development of a three-tier test to assess ninth grade students' misconceptions about simple electric circuits*. Unpublished master's thesis, Middle East Technical University, Ankara, Turkey.
- Peşman, H., & Eryılmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *The Journal of Educational Research*, 103(3), 208-222.
- Peterson, R. F., & Treagust, D. F. (1989). Grade-12 students' alternative conceptions of covalent bonding and structure. *Journal of Chemical Education*, 66(6), 459-460.
- Reynolds, A. J., & Walberg, H. J. (1992). A structural model of science achievement and attitude: An extension to high school. *Journal of Educational Psychology*, 84(3), 371-382.
- Şekercioğlu, A. G., & Kocakulah, M. S. (2008). Grade 10 Students' Misconceptions about Impulse and Momentum. *Journal of Turkish Science Education*, 5(2), 47-59.
- Şen, Ş. (2015). *Investigation of students' conceptual understanding of electrochemistry and self-regulated learning skills in process oriented guided inquiry learning environment* (Unpublished Dissertation). Hacettepe University, Ankara, Turkey.
- Şen, Ş. (2016). The relationship between secondary school students' self-regulated learning skills and chemistry achievement. *Journal of Baltic Science Education*, 15(3), 312-324.
- Şen, Ş., & Yılmaz, A. (2012). Effect of conceptual change texts assisted dual situated learning model on achievement. *Journal of Hacettepe University Education Faculty*, 42, 367-379.
- Şen, Ş., & Yılmaz, A. (2013). A phenomenographic study on chemical bonding. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, 7(2), 144-177.
- Sreenivasulu, B. & Subramaniam, R. (2013). University students' understanding of chemical thermodynamics. *International Journal of Science Education*, 35(4), 601-635.
- Taber, K. S. (2009). Challenging Misconceptions in the Chemistry Classroom: Resources to Support Teachers. *Educació química*, 4, 13-20.
- Taber, K. S. (2011). Models, molecules and misconceptions: A commentary on "secondary school students' misconceptions of covalent bonding". *Journal of Turkish Science Education*, 8(1), 3-18.
- Tan, D. K. C., & Treagust, D. F. (1999). Evaluating students' understanding of chemical bonding. *School Science Review*, 81(294), 75-84.
- Tan, K. C. D., Goh, N. K., Chia, L. S., & Treagust, D. F. (2002). Development and application of a two-tier multiple-choice diagnostic instrument to assess high school students' understanding of inorganic chemistry qualitative analysis. *Journal of Research in Science Teaching*, 39, 283-301.
- Temel, S., & Özcan, Ö. (2016). The analysis of prospective chemistry teachers' cognitive structure: the subject of covalent and ionic bonding. *Eurasia Journal of Mathematics, Science & Technology Education*, 12(8), 1953-1969.
- Thompson, F., & Logue, S. (2006). An exploration of common student misconceptions in science. *International Education Journal*, 7(4), 553-559.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159-169.
- Tsaparlis, G., & Papaphotis, G. (2002). Quantum-chemical concepts: Are they suitable for secondary students? *Chemistry Education Research and Practice*, 3(2), 129-144.
- Ünal, S., Coştu, B., & Ayas, A. (2010). Secondary school students' misconceptions of covalent bonding. *Journal of Turkish Science Education*, 7(2), 3-29.
- Uzuntiryaki, E., & Geban, Ö. (2005). Effect of conceptual change approach accompanied with concept mapping on understanding of solution concepts. *Instructional Science*, 33(4), 311-339.

- Voska, K. W., & Heikkinen, H. W. (2000). Identification and analysis of student conceptions used to solve chemical equilibrium problems. *Journal of Research in Science Teaching*, 37(2), 160-176.
- Walsh, W. B., & Betz, N. E. (2000). *Tests and assessments* (4th Ed.). Upper Saddle River, NJ: Prentice Hall.
- Wiersma, W., & Jurs, S. G. (1990). *Educational measurement and testing* (Second Ed.). Needham Heights, MA: Allyn and Bacon.
- Wuttirom, S., Sharma, M. D., Johnston, I. D., Chitaree, R., & Soankwan, C. (2009). Development and use of a conceptual survey in introductory quantum physics. *International Journal of Science Education*, 31(5), 631-654.
- Yayon, M., Mamlok-Naaman, R., & Fortus, D. (2012). Characterizing and representing student's conceptual knowledge of chemical bonding. *Chemistry Education Research and Practice*, 13, 248-267.
- Yürük, N. (2005). *An analysis of the nature of students' metaconceptual processes and the effectiveness of metaconceptual teaching practices on students' conceptual understanding of force and motion*. Unpublished Dissertation, Ohio State University, Ohio, USA.

From Editorial Board

This current issue, (Volume 14 Number 1), embraces seven articles of different areas in science education. These articles report important conclusions, some of which are outlined here. These are: (i) interactions between teacher and student, students' learning engagement and learning environment make a significant contribution to effective science learning; (ii) an improper relationship between school science and daily life language brings about misunderstanding of scientific language; (iii) pre-service science teachers' TPACK levels have a direct proportion with their grade level as well as their experiences with technology and teaching positively influence the TPACK levels; (iv) a systematic review of biology education research in Turkey from 1989 to 2015 reports that MS theses were higher than PhD ones before 2000s and most of the theses preferred "Quantitative" method as a research method; (v) The article examining middle school students' constructivist environment perceptions in Turkey addresses that students in small classes and rural schools were more positive in their classroom learning environments (i.e. thought provoking, collaboration, life relevance, concurrent learning and assessment, and bringing different viewpoints) than those who were in large classes and urban schools; (vi) the article entitled 'Effectiveness of the Proposed Training Formative Assessment Programme and its Impact on Teaching Style Improvements of Saudi Science Teachers in Saudi Arabia' reveals that science teachers adopted various strategies of formative assessment, sharing learning goals, open-ended questions, assessment-based evidence, and teacher–student dialogue; (vii) a three-tier concept test determining high school students' conceptual understanding of chemical bonding was developed and implemented to confirm its reliability and validity.

Happy reading!

On behalf of Editorial Board Members

Prof. Dr. Salih CEPNI

Editor-in-chief