



How Differences among Data Collectors are Reflected in the Reliability and Validity of Data Collected by Likert-Type Scales?

Mustafa Serdar KÖKSAL^a
İnönü University

Pelin ERTEKİN^b
İnönü University

Özgür Murat ÇOLAKOĞLU^c
Bülent Ecevit University

Abstract

The purpose of this study is to investigate association of data collectors' differences with the differences in reliability and validity of scores regarding affective variables (motivation toward science learning and science attitude) that are measured by Likert-type scales. Four researchers trained in data collection and seven science teachers who did not undergo any training, gathered data from 391 ninth-grade students. The data collection instruments were the "Motivation toward Science Learning Scale" and "Science Attitude Scale." Data collection applications were conducted in four stages, two of which were accomplished four weeks apart by the researchers. The remaining two stages were accomplished four weeks apart by the teachers. A principal component analysis, confirmatory factor analysis, Cronbach's alpha reliability analysis, Pearson correlation test for convergent validity, and t-test for the differences between the mean scores of each data collection stage were used for the data analysis. The results showed that motivation toward science learning and attitude toward science were high but the factor structures and reliability values, which were obtained by different data collectors, were different for the two scales. As another result, the convergent validity between the scores on the scales was shown to be sufficient for the measurements. However, the results of difference tests on the mean scores of the applications showed that there was a statistically significant difference between the mean scores of the two motivation scale applications by the teachers.

Keywords

Data Collector, Motivation toward Learning Science, Science Attitude, Validity, Reliability.

In science education literature, Likert-type scales are frequently used for data collection, but researchers prefer different data collectors when they carry out research using one type of scale. Although the same

scale is used in different studies, the use of different data collectors might make an important difference in the research results (Fraenkel & Wallen, 2003). The differences arising from data collectors are

- a **Mustafa Serdar KÖKSAL, Ph.D.**, is currently an associate professor of Science Education. Her research interests include the nature of science, epistemological beliefs, and gifted education. *Correspondence:* Inonu University, Faculty of Education, Department of Elementary Education, Malatya, Turkey. Email: bioeducator@gmail.com
- b **Pelin ERTEKİN, Ph.D.** student, is currently a research assistant of Science Education. Contact: Inonu University, Faculty of Education, Department of Elementary Education, Malatya, Turkey. Email: pelin.ertekin@inonu.edu.tr
- c **Özgür Murat ÇOLAKOĞLU, Ph.D.** student, is currently a research assistant of Computers and Educational Technologies. Contact: Bulent Ecevit University, Faculty of Education, Department of Computers and Educational Technologies, Zonguldak, Turkey. Email: ozgurcolakoglu@karaelmas.edu.tr

an important factor threatening internal validity in research (Fraenkel & Wallen, 2003). Therefore, data collector characteristics become an important factor in the data collection process (Fraenkel & Wallen, 2003; Miyazaki & Taylor, 2008). The scale implementation process includes procedures to take this into account and requires expertise. In this process, the implementers try to properly proceed using handbooks about the scale (Brener, McManus, Galuska, Lowry, & Wechsler, 2003). Undergoing training (or not) is an important component of data collection, but some of the studies in the field of science education do not give any information about data collectors (Akpınar, Aktamış, & Ergin, 2005; Gömleksiz & Bulut, 2006; Yıldız, Akpınar, Aydoğdu, & Ergin, 2006). Probably, data are frequently collected by teachers. However, how to develop and apply a scale for research is not taught to pre-service science teachers who are working toward their bachelor's degree at Turkish universities. In spite of the need for data collection to solve the problems in Turkey's educational system, there is no strong training course in line with this purpose. Turkey is among the least successful countries in the PISA examination (The Organisation for Economic Co-operation and Development, 2009), indicating a need to collect more data about where the problem lies. To meet this need, it is necessary to check the data collection process that use Likert scales for the data collector effect.

Although insufficient information on data collector characteristics is reported in papers, the differences among data collectors in terms of whether or not they have received training might change the reliability and validity of the scores collected by Likert scale applications. For example, Rogers (1976) stated that task- or individual-oriented data collection processes make a difference in consistency in data collection. Reliability and validity are characteristics of scores obtained from a scale and are two factors that have an effect on the quality of inference after the measurement (American Educational Research Association, 1999; Del Greco, Walop, & McCarthy, 1987). Discrepancies originating from the data collector can lead to differences in the values of reliability and validity, thereby negatively influencing the accuracy of inferences based on measurements. The importance of this problem in terms of obtaining results in survey research using Likert-type scales in science education sets the framework of this study. Thus, the problem is examined by investigating the reliability and validity of measurements regarding two affective variables (i.e., motivation toward

science learning and science attitude) that are measured using Likert scales in science education.

In education literature, motivation and attitude are frequently researched affective factors (Bong, 2001; Dede & Yaman, 2008; Douglas, 2006; Kahyaoglu, 2013; Koballa & Glynn, 2007; Oğuz Çakır, 2011; Osborne, Simon, & Collins, 2003; Pintrich, 1999; Pintrich & DeGroot, 1990; Savran & Çakıroğlu, 2001; Serin, 2009; Simpson, Koballa, Oliver, & Crawley, 1994; Temiz, 2010; Wigfield & Eccles, 2000; Yenice, Saydam, & Telli, 2012) that are measured with Likert-type scales (Çavaş, 2011; Dede & Yaman, 2008; Savran & Çakıroğlu, 2001; Tuan, Chin, & Shieh, 2005; Yılmaz & Çavaş Huyugüzel, 2007; Yumuşak, Sungur, & Çakıroğlu, 2007). Motivation is an affective characteristic that is effective on acting for reaching a purpose (Brophy, 1998). For research on motivation in science education, "Students' Motivation toward Science Learning (SMTSL)" developed by Tuan et al. (2005) is an important scale because it has been applied to large samples and has high values of reliability and validity. Moreover, this scale was adapted to Turkish by Yılmaz and Çavaş Huyugüzel (2007). On the other hand, the "Science Attitude Scale (SAS)" developed by Geban, Ertepinar, Yılmaz, Atlan, and Şahpaz (1994) is another Likert-type scale used frequently in Turkey (Bilgin & Karaduman, 2005; Çavaş, 2011; Kenar & Balci, 2012; Özyılmaz & Hamurcu, 2005; Tatar & Kuru, 2009; Ünal & Ergin, 2006). Both of these affective focused studies present information about the reliability and validity values, but no information is given about data collectors. Consequently, investigating the possible effect of data collector differences on validity and reliability is an important contribution for current science education studies and future studies that will use Likert-type scales.

The purpose of this study is to investigate how data collector differences are reflected in the reliability and validity of scores regarding affective variables (motivation toward science learning and science attitude) that are measured by Likert scales.

Method

In this study, reliability and validity values of the data gathered by different data collector groups were investigated by utilizing a survey approach (Karasar, 1999; Wallen & Fraenkel, 2001). The data were collected from 391 (184 female, 107 male) ninth-grade Anatolian high school students. The data collectors were four researchers (2 female, 2

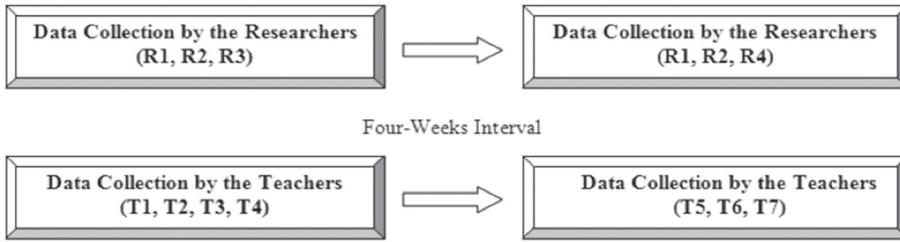


Figure 1: Data collection applications.

male) and seven science teachers (1 female, 6 male). The researchers received a two-week training on how to apply the scale (two hours per week). The training content consisted of introducing the research subject of the scale application, explaining the purpose of the application, stating possible advantages and disadvantages of the application, and explaining ethical subjects, dress style, and use of language. However, the teachers applied the scales without undergoing any training. In this study, the applications of the four data collection processes were conducted separately. These applications included two beginning scale applications and two scale applications implemented four weeks later. Figure 1 shows the model of the scale applications.

The data collection instruments were the SMTSL and SAS. The SMTSL, the original of the first scale, was developed by Tuan et al. (2005). Then the scale was adapted into Turkish by Yılmaz and Çavaş Huyugüzel (2007). The Turkish version of the scale consists of six factors (self-efficacy, active learning strategies, science learning value, performance goal, achievement goal, and learning environment stimulation) and includes 33 items. The result of the reliability analysis of the scores showed that Cronbach's alpha values of the factors were between .54 and .85; on the other hand, the reliability analysis of the total scores on the scale was .87. Two examples of the scale items are "When I find the science content difficult, I do not try to learn it" and "In science, I think that it is important to learn to solve problems." The SAS, the second scale, was developed by Geban et al. (1994). Başer (1996) reported that the SAS included 15 items and had one factor. In addition, Cronbach's alpha value of the scores on the scale was .83. Two example items in the scale are "I am bored when I study science subjects" and "I want to learn more about science subjects." Confirmatory and explanatory factor analyses (*principal component analysis and varimax rotation*) for construct validity, Cronbach's alpha reliability analysis, Pearson correlation test for convergent validity, and a *t*-test for the differences between mean scores of each data collection stage

were used for the data analysis. For the *t*-test analysis, *Bonferroni correction* was done, and the alpha value was determined as .006. AMOS and SPSS 18 package programs were used for all analyses.

Findings

The findings of this study are presented under three main headings: construct validity and reliability, convergent validity, and a *t*-test for the differences between the mean scores in each application.

Construct Validity and Reliability

Findings Regarding Construct Validity and Reliability (SMTSL): The confirmatory factor analysis results for each application indicated that although (X^2/sd) was between 1.58 and 2.28, the other indexes for each application were not acceptable for the proposed factor model (GFI: .62-.71; CFI: .64-.77; RMSEA: .08-.11) (Hoyle, 2000; Marsh, Balla, & McDonald, 1988; Marsh & Hocevar, 1988; Raykov & Marcoulides, 2006). On the other hand, (X^2/sd) and RMSEA indexes showed differences between the trained and untrained data collectors in terms of focused variables.

Because of the confirmatory analysis results, an explanatory factor analysis was carried out. Before the analysis, the Kaiser-Meyer-Olkin (KMO) measure of sample adequacy and Barlett's test of sphericity values were calculated. The results ($KMO > .60, p < .05$) showed that the data were suitable for factor analysis (Sharma, 1996; Tavşancıl, 2002). According to the principal component analysis, the scores collected by different data collectors revealed different factor structures and also explained that total variances are different for each application. The item loading values and loaded factors for each application were also different for the same instrument. The reliability results showed that the reliability value of each factor was quite different and between .34 and .86. The total reliability value for each data collector group was .82 and .89, respectively.

Findings Regarding Construct Validity and Reliability (SAS): The confirmatory factor analysis results for each application indicated that except for “ X^2/sd ” (1.40–3.45) and CFI value (.83-.95), the other indexes for each application are not acceptable for the proposed factor model (GFI: .77-.82; RMSEA: .08-.14) (Hoyle, 2000; Marsh, Balla, & McDonald, 1988; Marsh & Hocevar, 1988; Raykov & Marcoulides, 2006).

After the confirmatory factor analysis, it was decided to carry out an explanatory factor analysis. According to the KMO measure of sample adequacy and Bartlett’s test of sphericity values ($KMO > .60$, $p < .05$), the data were suitable for factor analysis (Sharma, 1996; Tavşancıl, 2002). When the results of the principal component analysis were investigated, similar results with SMTSL were found, and also, the omitted items after analyses were not common for each application. The reliability of each factor was quite different, ranging from .60 to .92. The total reliability value for each data collector was .88 and .92, respectively.

Convergent Validity

The convergent validity was examined by investigating correlations between the scores on the motivation and attitude scales. The results indicated that there was a statistically significant and positive relationship between the scores of motivation and attitude for each application ($r = .56-.66$, $p < .05$). These results showed the expected results in terms of convergent validity (Singh, Graville, & Dika, 2002; Tuan et al., 2005).

T-Test for the Differences between the Mean Scores of Each Application

The results of difference tests between the mean scores of the applications showed that there was not any statistically significant difference between the mean scores of the two motivation and attitude scale applications by the researchers ($t_{SMTSL} = 1.66$, $p > .006$; $t_{SAS} = 0.45$, $p > .006$). Therefore, there was no practical importance of the results in terms of effect sizes (Coe, 2002).

The non-significant difference between the mean scores of the two attitude scale applications by teachers was also determined ($t_{SAS} = 0.51$, $p > .006$). However, there was a statistically significant difference between the mean scores of the two motivation scale applications by the teachers ($Z = 3.15$, $p < .006$). Consequently, there was not

any statistically significant difference between the mean scores of the researchers’ and teachers’ first and second motivation, as well as attitude scale applications (first application: $t_{SMTSL} = .39$, $p > .006$; $t_{SAS} = 1.09$, $p > .006$; second application: $t_{SMTSL} = .2.59$, $p > .006$; $t_{SAS} = 0.95$, $p > .006$).

Discussion and Suggestions

This study found that reliability and validity values differed significantly across the data collection applications. According to the confirmatory and explanatory factor analyses, factor structures, items loadings in the factors, and index values differed between the two applications conducted four weeks apart by the researcher and teacher. These differences may arise from differences in the data collectors’ characteristics despite their having had the same training (Fraenkel & Wallen, 2003). Especially, definite differences were seen among the applications of the teacher data collectors. For instance, an important difference is statistically significant between the two applications by the teachers. These findings show that motivation data obtained by teachers yielded two different results when collected at different times. Consequently, it can be speculated that this is a reflection of differences in the data collectors for construct validity and data stability of SMTSL. Total reliability values regarding the motivation scales were similar between the data of researchers and teachers. However, the important point is that the reliability of different factors cannot be compared because the factors do not share a common structure.

A look at the second variable of this study shows that attitude scores are different in terms of factor structures and reliability values. Attitude is suggested as a susceptible affective variable of data collector characteristics by Pol and Ponzurick (1989). Therefore, the findings of this study also supported Pol and Ponzurick’s suggestion. These findings mean that motivation, similar to attitude, is susceptible to data collector characteristics. Also, these findings supported previous studies by Eryılmaz (2002), Behi and Nolan (1996), and Miyazaki and Taylor (2008), who explained that training on the data collection process, experience in data collection, gender, race, and age were important factors in explaining differences in the data collected by various data collectors. Moreover, Sondergeld and Johnson (2014) emphasized that factor structures in scales may differ depending on the sample, and this creates difficulties in the comparison of different study results. Thus, it can

be considered that researchers' ignorance of data collector differences threatens the reliability and validity of data obtained from Likert-type scales. An important way to reduce these differences is to train the data collectors, but this study's findings showed that training status alone is not sufficient to provide strong reliability and validity.

The other finding regarding convergent validity supported the literature in terms of the relationship between motivation and attitude (Singh et al., 2002; Tuan et al., 2005). In all of the applications, there

was a statistically significant relationship between motivation and attitude. Therefore, this result showed that the measurements had convergent validity.

Based on the findings of this study, it was suggested that data collector characteristics should be taken into account when Likert-type instruments are used to collect data on motivation and attitude. At the same time, other affective variables such as self-efficacy and anxiety should also be examined using a similar approach.

References/Kaynakça

- Akpınar, E., Aktamış, H. ve Ergin, Ö. (2005). Fen bilgisi dersinde eğitim teknolojisi kullanılmasına ilişkin öğrenci görüşleri. *The Turkish Online Journal of Educational Technology*, 4(1), 93-100.
- American Educational Research Association. (1999). *Standards for educational and psychological testing*. Washington, DC: National Council on Measurement in Education.
- Başer, M. (1996). *Kavram değiştirme yönteminin öğrencilerin ısı ve sıcaklık kavramlarını anlamalarına ve fen tutumlarına etkisi* (Yüksek lisans tezi, Orta Doğu Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Ankara). <https://tez.yok.gov.tr/UlusalTezMerkezi> adresinden edinilmiştir.
- Behi, R., & Nolan, M. (1996). Causality and control: Threats to internal validity. *British Journal of Nursing*, 5(6), 374-377.
- Bilgin, İ. ve Karaduman, A. (2005). İşbirlikli öğrenmenin 8. sınıf öğrencilerinin fen dersine karşı tutumlarına etkisinin incelenmesi. *İlköğretim Online*, 4(2), 32-45.
- Bong, M. (2001). Role of self-efficacy and task-value in predicting college students' course performance and future enrollment intentions. *Contemporary Educational Psychology*, 26, 553-570.
- Brener, N. D., McManus, T., Galuska, D. A., Lowry, R., & Wechsler, H. (2003). Reliability and validity of self-reported height and weight among high school students. *Journal of Adolescent Health*, 32(4), 281-287.
- Brophy, J. (1998). *Motivating students to learn*. Madison, WI: McGraw Hill.
- Cavaş, P. (2011). Factors affecting the motivation of Turkish primary students for science learning. *Science Education International*, 22(1), 31-42.
- Coe, R. (2002). *It's the effect size, stupid: What effect size is and why it is important*. Retrieved from <http://www.leeds.ac.uk/educol/documents/00002182.html>
- Dede, Y. ve Yaman, S. (2008). Fen Öğrenmeye Yönelik Motivasyon Ölçeği: Geçerlik ve güvenilirlik çalışması. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED)*, 2(1), 19-37.
- Del Greco, L., Walop, W., & McCarthy, R. H. (1987). Questionnaire development: 2. validity and reliability. *CMAJ: Canadian Medical Association Journal*, 136(7), 699-700.
- Douglas, L. (2006). Motivational factors, learning strategies and resource management as predictors of course grades. *College Student Journal*, 40(2), 423-428.
- Eryılmaz, A. (2002). Effects of conceptual assignments and conceptual change discussions on students' misconceptions and achievement regarding force and motion. *Journal of Research in Science Teaching*, 39(10), 1001-1015.
- Fraenkel, J. R., & Wallen, N. E. (2003). *How to design and evaluate research in education* (5th ed.). New York: McGraw-Hill Publishing Co.
- Geban, Ö., Ertepinar, H., Yılmaz, G., Atlan, A. ve Şahpaz, Ö. (1994, Eylül). *Bilgisayar destekli eğitimin öğrencilerin fen bilgisi başarılarına ve fen bilgisi ilgilerine etkisi*. I. Ulusal Fen Bilimleri Eğitimi Sempozyumu'nda sunulan bildiri, Dokuz Eylül Üniversitesi, Buca Eğitim Fakültesi, İzmir.
- Gömlüksiz, M. N., & Bulut, I. (2006). An evaluation of elementary school students' opinions toward science classes. *Eurasian Journal of Educational Research*, 23, 106-116.
- Hoyle, R. (2000). Confirmatory factor analysis. In H. E. A. Tinsley, & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 465-497). California: Academic Press.
- Kahyaoglu, M. (2013). A comparison between gifted students and non-gifted students' learning styles and their motivation styles towards science learning. *Educational Research and Reviews*, 8(12), 890-896.
- Karasar, N. (1999). *Bilimsel araştırma yöntemi: Kavramlar, ilkeler, teknikler*. Ankara: 3A Araştırma Eğitim Danışmanlık Ltd.
- Kenar, İ. ve Balcı, M. (2012). Fen ve Teknolojiye Yönelik Tutum Ölçeği: İlköğretim 4 ve 5. sınıf örneği. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 34, 201-210.
- Koballa, T. R., & Glynn, S. M. (2007). *Attitudinal and motivational constructs in science learning*. In S. Abell & N. Lederman (Eds.), *Handbook of research on science education* (pp. 75-102). Mahwah, NJ: Lawrence Erlbaum.
- Köksal, M. S. (2012). Adaptation study of motivation toward science learning questionnaire for academically advanced science students. *Chemistry: Bulgarian Journal of Science Education*, 21(1), 29-44.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391-410.
- Marsh, H. W., & Hocevar, D. (1988). A new more powerful approach to multitrait-multimethod analysis: Application of second-order confirmatory analysis. *Journal of Applied Psychology*, 73, 107-117.
- Miyazaki, A. D., & Taylor, K. A. (2008). Researcher interaction biases and business ethics research: Respondent reactions to researcher characteristics. *Journal of Business Ethics*, 81, 779-795.
- Organisation for Economic Co-operation and Development. (2009). *Take the test: Sample questions from OECD's PISA assessments*. Paris: Author.
- Oğuz Çakır, B. Z. (2011). *The influence of argumentation based instruction on sixth grade students' attitudes toward science, conceptual understandings of physical and chemical change topic and argumentativeness* (Master's thesis, Middle East Technical University, Department of elementary science and mathematics education, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), 1049-1079.
- Özyılmaz, A. G. ve Hamurcu, H. (2005). Çoklu zekâ kuramı tabanlı öğretimin öğrencilerin fen başarısı, tutumları ve hatırdı tutma üzerindeki etkileri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 28, 178-187.
- Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31, 459-470.
- Pintrich, P. R., & De Groot E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33-50.
- Pol, L. G., & Ponzurick, T. G. (1989). Gender of interviewer/gender of respondent bias in telephone surveys. *Applied Marketing Research*, 29, 9-13.
- Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). London: Lawrence Erlbaum Associates, Publishers.
- Rogers, T. F. (1976). Interviews by telephone and in person: Quality of responses and field performance. *Public Opinion Quarterly*, 40, 51-65.
- Savran, A., & Çakıroğlu, J. (2001). Preservice biology teachers perceived efficacy beliefs in teaching biology. *Hacettepe University Journal of Education*, 21, 105-112.

- Serin, G. (2009). *The effect of problem based learning instruction on 7th grade students' science achievement, attitude toward science and scientific process skills* (Doctoral dissertation, Middle East Technical University, Department of Secondary Science and Mathematics Education, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Sharma, S. (1996). *Applied multivariate techniques*. New York: John Wiley.
- Simpson, R. D., Koballa, T. R., Oliver, J. S., & Crawley, F. E. III. (1994). Research on the affective dimension of science learning. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 542–558). New York: Macmillan.
- Singh, K., Graville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interests and academic engagement. *Journal of Educational Research*, 95(6), 323–332.
- Sondergeld, T. A., & Johnson, C. A. (2014). Using Rasch measurement for the development and use of affective assessments in science education research. *Science Education*, 98(4), 581–613.
- Tatar, N., & Kuru, M. (2009). Inquiry-based learning approach versus descriptive methods: Effects on elementary students' attitudes towards science. *Pamukkale Eğitim Fakültesi Dergisi*, 25, 142-152.
- Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi [Measurement of attitudes and data analysis with SPSS]*. Ankara: Nobel Yayınları.
- Temiz, B. (2010). İlköğretim 6. sınıf öğrencilerinin "Vücudumuzda Sistemler" ünitesindeki akademik başarı ve fene karşı tutumlarına örnek olay destekli 5E öğretim modelinin etkisi (Yüksek lisans tezi, Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara). <https://tez.yok.gov.tr/UlusalTezMerkezi> adresinden edinilmiştir.
- Tuan, H., Chin, C., & Shieh, S. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International Journal of Science Education*, 27(6), 634-659.
- Ünal, G. ve Ergin, Ö. (2006). Buluş yoluyla fen öğretiminin öğrencilerin akademik başarılarına, öğrenme yaklaşımlarına ve tutumlarına etkisi. *Türk Fen Eğitimi Dergisi*, 3(1), 36-52.
- Wallen, N. E., & Fraenkel, J. R. (2001). *Educational research: A guide to the process*. Mahwah, NJ: Lawrence Erlbaum.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68-81.
- Yenice, N., Saydam, G. ve Telli, S. (2012). İlköğretim öğrencilerinin fen öğrenmeye yönelik motivasyonlarını etkileyen faktörlerin belirlenmesi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi (KEFAD)*, 13(2), 231-247
- Yıldız, E., Akpınar, E., Aydoğdu, B. ve Ergin, Ö. (2006). Fen bilgisi öğretmenlerinin fen deneylerinin amaçlarına yönelik tutumları. *Türk Fen Eğitimi Dergisi*, 3(2), 2-18.
- Yılmaz, H., & Çavaş Huyugüzel, P. (2007). Reliability and validity study of the Students' Motivation toward Science Learning (SMTSL) Questionnaire. *Elementary Education Online*, 6(3), 430-440.
- Yumuşak, N., Sungur, S., & Çakıroğlu, J. (2007). Turkish high school students' biology achievement in relation to academic self-regulation. *Educational Research and Evaluation*, 13, 53-69.