An Investigation of Feedback Strategies Used by Science Teachers in the Classroom Setting: A Mixed-Methods Research

Ufuk Özkale [i]
Ministry of National Education

Sedat Kanadlı [ii]
Mersin University

Abstract

This study classifies the verbal feedback of science teachers into praise statements, effort-based feedback statements, negative feedback statements, and ability-based statements, which are also regarded as feedback strategies. The study aims to investigate the feedback strategies used by science teachers in the classroom setting using descriptive research with a general survey design within an exploratory sequential design of mixed-methods research. It employs the Science Course Feedback Perception Scale in data collection. The population of the study consists of the 6th, 7th, and 8th-grade students (N = 1696) and secondary school science teachers (N = 51) affiliated to the Turkish Ministry of National Education in the central districts of Mersin, Turkey in the 2016-2017 and 2017-2018 school years. The sample is formed through a theoretical sampling method for qualitative data and convenience sampling and cluster sampling methods for quantitative data. The qualitative data are collected by open-ended questions designed based on expert opinion. The items are reviewed based on expert opinion and the content validity index of the scale is examined. Following that, exploratory factor analysis is performed to test the construct validity of the scale. The factor analysis identified a 4-factor model with 34 items explaining 55.73% of the total variance of the scale. Then, the 4-factor structure of the scale is tested by confirmatory factor analysis. Lastly, the types of feedback given by teachers to the students in the science course are analyzed in terms of gender and grade level.

Keywords: Feedback Strategies, Feedback Types, Science Course, Teacher Feedbacks

DOI: 10.29329/ijpe.2021.329.28

---

[i] **Ufuk Özkale,** Expert, Elementary School Teacher, Ministry of National Education, ORCID: 0000-0002-8121-5470

Correspondence: ozkaleufuk@gmail.com

[ii] **Sedat Kanadlı,** Assoc. Prof. Dr., Educational Sciences, Faculty of Education, Mersin University, ORCID: 0000-0002-0905-8677

INTRODUCTION

Many factors influence the level of learning of students in the classroom environment. These factors range from student readiness, attitude towards the course, and intelligence level to environmental factors such as light, sound, color, and class size in the classroom context. Taking the factors affecting student learning into consideration, teacher (30%) has the largest effect after genetic factors (50%; Hattie, 2003). According to Hattie (2003), this effect is created by teachers who have deep pedagogical content knowledge, have high expectations for students, build positive teacher-student relationships, monitor their students, and provide feedback. Feedback alone has the strongest effect among these teacher qualifications, and as per Cohen's *d* (*d* = 0.73), it has a moderate effect (Thalheimer & Cook, 2002).

There are many definitions of feedback in the literature. However, most comprehensively, feedback is defined as particular information provided for improving student performance considering the result obtained from the comparison between observed and standard performance (Van de Ridder, Stokking, McGaghie, & Ten Cate, 2008). Put differently, the teacher-provided feedback refers to student's current knowledge and understanding and the difference between the expectations concerning this knowledge and understanding (Orlich, Harder, Callahan, Trevisan, & Brown, 2007, p.337). Feedback plays an important role in the development of students, as it provides the knowledge necessary to inform them regarding their performance and development in terms of learning and teaching (Sadler, 1989; Higgins, 2000; Taras, 2005). Additionally, in his meta-analysis study, Walberge (1984) reported that feedback ranks third among the 26 factors affecting student achievement. Based on his review of 95 studies, Walberg argues that the feedback given during the teaching process increases student achievement from 50 to 89% (cited in Adrienne, 1997). Feedback has a significant effect on cognitive, affective, and motivational processes as well as influences student's self-understanding (self-confidence and control; Vollmeyer & Rheinberg, 2005). That is why the quality of feedback the student receives is important. That is to say, the message of feedback, the timing of feedback, the place of feedback, the reasons for getting and giving feedback, and in what ways the feedback is received are important in terms of its effectiveness (Brinko, 1990).

The study of feedback from different aspects has brought about many classifications in the literature along with different definitions despite they share common features related to feedback. Therefore, many classifications related to feedback are available. According to Burnett (1996), teacher feedback is divided into four: *(i)* Negative feedback is to tell students that their current performance is faulty or is not at the desired level (e. g. what you have done is not enough). *(ii)* Positive feedback comprises statements of praise, appreciation, and confirmation, telling students that their current performance is at the desired level (e.g. Good Job! Nice work). *(iii)* Effort-based feedback points out to the student's efforts or desires in the process of showing up the desired performance or things the student needs to do to achieve the desired performance (e.g. You are a real hard-worker). *(iv)* Ability-based feedback comprises statements that emphasize academic or individual abilities, skills, and other characteristics in students' performance in the relevant course that distinguish them from other students (e.g. you are really good at this). Yet, Burnett (1996), in his study, asserted that statements students hear from people around them are effective in their achievement and that praise statements contribute more to their achievement than do the negative ones. Also, giving verbal feedback is stated to be proportionally more effective than giving the written one (Edmonds, 1978; cited in Dökmen, 1982).

This study aims to determine the feedback strategies used by science teachers in the classroom setting. The feedback strategies in this study were defined under the four feedback types of "praise statements", negative feedback statements", "effort-based feedback statements", and "ability-based feedback statements" that are verbally used in the classroom by science teachers. To this end, Burnett's conceptual framework was used for feedback strategies. In addition to this conceptual construct, the praise statements heard by students during the class time were added, too. The term "praise" comes from the Latin origin of "pretiare", meaning to be of high value (Shepell, 2000) which

also includes appreciation, admiration, and approval of a person's value. Teacher praise has a positive effect and is a more intense and detailed response in student behavior than is feedback (Blote, 1995). According to research by Hitz and Driscoll (1989) on teacher's use of praise in the classroom, effective praise is thought to occur when the teacher positively approves the work of students.

Moreover, there is a limited number of research on feedback strategies used by teachers in national literature and the existing research is mostly conducted using qualitative research methods (Çobakçor, Akşan, Öztürk, & Çimer, 2011; Türkdoğan & Baki, 2012; Köğce, 2012; Cengiz, 2012). As such, no study was found utilizing a measurement tool related to feedback strategies and study with a large sample group. Thus, a measurement tool is needed in the national literature to work on a large group of students and to determine the feedback strategies of science teachers. In this study, a scale was developed to determine the feedback strategies used by science teachers during the class. So this study is thought to be of use to researchers in studies yet to be done in the field and to guide future professional development activities organized for promoting feedback competences of teachers.

Thus, this study sought answer to the question "What are feedback strategies used by science teachers in the classroom setting, and do these strategies significantly differ in terms of gender and grade level? According to this general purpose, the sub-problems of the study were determined as in the following:

1- What are the feedback scores of students (praise score, effort-based feedback score, negative feedback score, and ability-based feedback score)?

2- Do the feedback scores of students significantly differ by gender?

3- Do the feedback scores of students significantly differ by grade level?

METHOD

This study is a mixed-methods research. Mixed-methods research is defined as a study where the researcher collects and analyzes data, includes the findings, and makes inferences by employing qualitative and quantitative approaches or methods in a single study or research program (Tashakkori & Creswell, 2007). Of mixed-methods designs, the exploratory sequential design was used in this study. Meanwhile, the exploratory sequential design is a two-stage sequential design in that the researcher begins exploring the subject qualitatively prior to the quantitative research. In many applications of this chaining design, the researcher develops a tool as a middle step between the stages created based on the qualitative results and uses this tool when collecting quantitative data (Creswell, Fetters, & Ivankova, 2004). The scale items were identified by primarily conducting a qualitative study to determine the perceived feedback of students followed by scale development studies, and, lastly, the quantitative study responded to the research questions. The research procedure is illustrated in Figure 1.

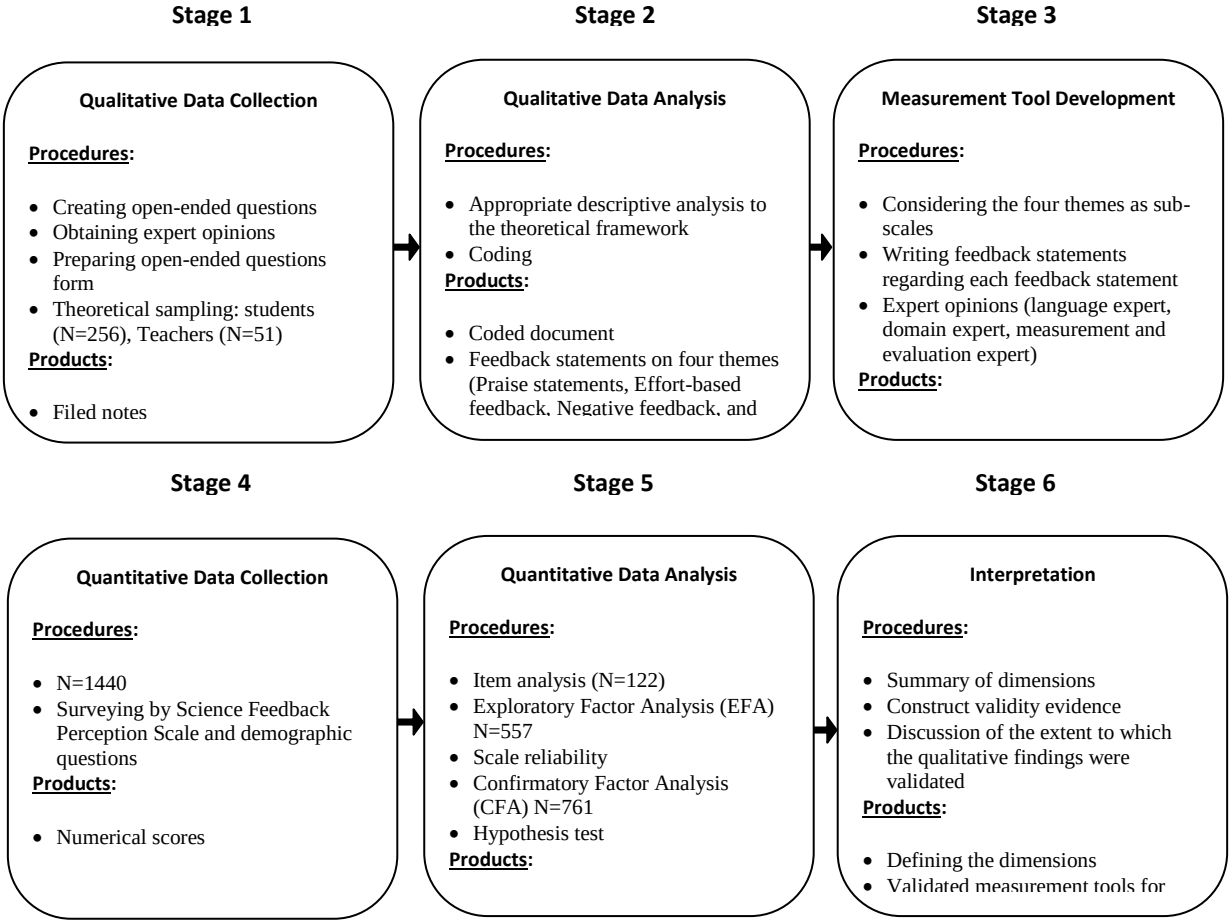

Figure 1. Stages of research

Study group

The study group is consisted of three groups according to the peculiarity of exploratory design in this study. The first group was formed to determine the scale items, the second group to determine the psychometric characteristics of the scale, and the third group to answer the research questions. The first group consisted of secondary school students and teachers in the central districts of Mersin province. Students were 256 persons in total studying in grade 6, 7, and 8. Of these students, 46.87% (n = 120) were females and 53.13% (n = 136) were males. Besides, 33.98% (n = 87) of students were studying in grade 6, 33.21% (n = 85) in grade 7, and 32.81% (n = 84) in grade 8. Additionally, there were 51 science teachers, of whom 54.90% (n=28) were females and 45.10% (n=23) males.

The second sample group was determined through a convenience sampling method for item analysis and exploratory factor analysis of the scale, which included students from grades 6, 7, and 8, studying in central districts of Mersin province. Item analyses were carried out with 122 students of grades 6, 7, and 8 (50.80% female; 49.20% male; age M = 12.85; SD = 0.91). Moreover, the Exploratory Factor Analysis (EFA) was conducted based on data from 557 grades 6, 7, and 8 students (55.66% female; 44.34 male; age M = 12.89; SD = 0.94) to investigate the factor structure of the scale.

The third study group was determined a disproportionate cluster sampling method to examine the model fit indices of the scale and to perform descriptive statistical analyses. For this purpose, the sample included 761 students (seven secondary schools) from 54.620 (study universe) secondary school students in grades 6, 7, and 8 in central districts of Mersin province in the 2017-2018 school year by considering a deviation of 0.05 and reliability level of α=0.01. Of these students, 57.42% were females and 42.58% males (age M = 12.47; SD = 1.07). Of them, 36.27% (276) were of grade 6, 32.32% (246) of grade 7, and 31.40% (239) of grade 8.

Data collection tools

In the qualitative part of the study, forms of six open-ended questions to teachers and five to students were administered to determine the "praise statements", "effort-based feedback statements", "negative feedback statements", and "ability-based feedback statements" that students say have heard from their science teachers and that teachers say they have uttered to students in the science class. After devising the questions of the form content, opinions of three experts, including a measurement and evaluation expert, a domain expert, and a curriculum development expert, were obtained. Besides, the opinions of a language expert were sought to check the meaning and spelling of the questions. After necessary corrections, these forms were administered to the teachers and students.

The "Science Course Feedback Perception Scale", consisting of feedback statements that emerged after qualitative data analysis of responses given to open-ended questions, was developed. In the quantitative part, however, this scale was employed to collect data on feedback statements that students hear from their teachers in the classroom.

Data analysis

In the qualitative part of the study, the data were analyzed using descriptive (deductive) content analysis, and feedback statements grouped under the themes fitting the theoretical framework were determined. Miles and Huberman's (1994) formula was applied to ensure inter-coder reliability between these coding. Content validity indexes were computed by having seven experts to rate the feedback statements by obtaining expert opinions, and excluding irrelevant items. This way, the final form was obtained before the exploratory factor analysis. Exploratory factor analysis, Mahalanobis distance test for outliers, multicollinearity and singularity tests, item analysis, confirmatory factor analysis, and split-half reliability tests were conducted using relevant package programs in the analysis of quantitative data analysis. In later stages of the analyses, percentages, frequency tables, normality tests, and other normality criteria were analyzed. Since group sizes were 50 and above, the Kolmogorov-Smirnov test of normality was employed to test the normal distribution of the data (Büyüköztürk, 2015, p.42). However, looking at the test results alone is not enough for the normality of distribution. In addition to normality tests, the decision on the normality of distribution was made examining the Histogram, Q-Q Graph, Skewness and Kurtosis values, and the values obtained by dividing the Skewness and Kurtosis values by their respective standard errors. Therefore, when deciding, the skewness and kurtosis values ranging between -1.50 and +1.50 (Tabachnick & Fidell, 2013) and the values ranging between -3.00 and +3.00 that are obtained by dividing the skewness and kurtosis values by their respective standard errors were taken into consideration (Kline, 2011; p. 63). Attention was also paid when examining the graphs that the points on the Q-Q graph are on or near the 45-degree line, the lines at the top and bottom of the box in P-P graph are close to each other, and the horizontal line in the box aligns to the center of the box (Morgan et al., 2004, p.60). When the normality conditions were satisfied, t-test and One-Way Variance Analysis (One-Way ANOVA) and, when not, Man-Whitney U and Kruskal Wallis tests were applied.

FINDINGS

Findings from the qualitative data analysis

Primarily, a descriptive analysis of responses given to the open-ended questions was conducted in this study (Strauss & Corbin, 1990). The data in this method, also called deductive analysis, are analyzed according to the existing frameworks (theoretical or conceptual), themes, and codes (Patton, 2002, p.483). Table 1 provides a theoretical framework for interpreting the data.

Table 1. Codes and definitions of the theoretical framework used in qualitative data analysis

Codes	Code Definitions	Sample Statements
Praise Statements	Unspecific performance-approving praise statements of teachers regarding the performance of students in the class without going into details.	Yeah, Ok, Right, You got a plus (+)! Good Job! Way to go! You are great! Congratulations! I am proud of you.
Negative Feedback Statements	Negative statements of teachers with devastating and threatening content regarding the personality, character, and performance of the students, rather than their performance in the course.	Don't you get it yet? What is there not to understand in this? Lazy! You got a minus (-).
Effort-Based Feedback Statements	Teacher's statements pointing out the student's efforts and desires in the process of showing up the desired performance or things the student needs to do to achieve the desired performance.	Let's do this question together?/ You should do it again. You need to work harder. / Think well, you can do it. /This is not you. You should get over yourself.
Ability-Based Feedback Statements	Teacher statements on the individual ability of students and the meta-cognitive performance or academic abilities they use when demonstrating this performance, apart from the performance students show in the course.	Just a question of your type! You are very good at science. Einstein and Newton have also answered this question as you did.

Validity and Reliability

To determine the inter-coder reliability for the quantitative part of the study, 30 forms of given responses to open-ended questions, i.e. 25 student and five teacher forms, were randomly selected and given to another coder. Coder reliability was determined by the formula [agreed codes / (agreed + disagreed codes)] x 100 proposed by Miles and Huberman (1994). The reliability values for each type of feedback are given in Table 2.

Table 2. Miles and Huberman reliability values of feedback types for teachers and students

Codes	Miles and Huberman (MH) Reliability Formula Values
Praise Statements	MH: 55 / (55+8) = 0.87
Negative Feedback Statements	MH: 63 / (63+10) = 0.86
Effort-Based Feedback Statements	MH: 68 / (68+13) = 0.83
Ability-Based Feedback Statements	MH: 43 / (43+3) = 0.93

Two coders agreed on 55 codes for praise statements while they did not agree on eight others. Accordingly, the inter-coder reliability was calculated as 87%. They agreed on 63 codes for negative feedback statements but did not on ten codes. Hence, the inter-coder reliability was calculated as 86%. For effort-based feedback statements, they agreed on 68 codes while they did not on 13. Accordingly, the inter-coder reliability was calculated as 83%. Lastly, for ability-based feedback statements, they agreed on 43 codes, while they did not on 3. Therefore, the inter-coder reliability was calculated as 93%. The consensus between coders must range between 80-90% (Miles & Huberman, 2015, p.64). Accordingly, valid reliability was ensured between the coders.

Findings on the scale development process

Content validity index (CVI)

Expert opinion forms were prepared concerning the statements in the item pool of 66 statements and the content validity indexes of these forms were calculated before performing the exploratory factor analysis. The content validity index is a method that is used for evaluating the expert opinions obtained for the scale items. In this method, first of all, the items in the item pool were scaled in a Likert format like "Completely Relevant", "Quite Relevant", "Somewhat Relevant", and "Irrelevant" and the expert opinion form was prepared. Seven experts were identified for the assessment of the items. One of these experts was a measurement and evaluation expert and three were curriculum development experts. Others were the domain experts. The responses given to the items by experts were rated as "1" if they were "Completely Relevant" or "Quite Relevant" and "0" if they were responded as "Somewhat Relevant" or "Irrelevant".

First of all, the item-level content validity indexes (I-CVI) were examined. In order to eliminate the chance factors when evaluating these indexes, if there were five or fewer than five experts, then all the experts should have rated an item as "1" to be accepted. However, if the number of experts is seven, at least two experts giving a score of zero will suffice to eliminate an item. The scores obtained from the experts were averaged for each item. For instance, the I-CVI for item1 and other items was calculated as "(1+1+1+1+0+1+1)/Number of Experts". Lynn (1986) suggests that the I-CVI value should not be smaller than 0.78. Therefore, of 66 items in the expert review form, 14 items were excluded in that they were under this cut-off value. After items with an I-CVI value of below 0.78 were excluded, the scale-level content validity indexes (S-CVI) were examined. S-CVI is the value which is obtained by dividing the total number of items rated "1" as a result of expert review by the sum of all items. After calculating this for each expert, the arithmetic mean of these values gives us the S-CVI value. According to Sauls (2004), S-CVI should not be less than 0.81. This value was 0.92 in the current study. It suggests that feedback statements are of high validity in terms of the content.

Exploratory factor analysis (EFA)

Exploratory Factor Analysis (EFA) and Cronbach's Alpha internal consistency analysis were performed using relevant package programs to determine the psychometric properties of the Science Course Feedback Perception Scale followed by a Confirmatory Factor Analysis (CFA) using a relevant statistical package program. For EFA, first of all, the assumptions of sample size, missing data, outlier analysis, multivariate normality and linearity, multicollinearity and singularity, and lastly the factorability of R were ensured (Tabachnick & Fidell, 2007).

It is worthwhile to note that the size of the sample group on which the data collection is performed is significant in scale development and adaptation studies. Therefore, 750 individuals reached in this study, and this size adequate for factor analysis (Comrey & Lee, 1992; Field, 2000). Item-level missing values were identified in the dataset obtained from the application and, thereby, 34 (n = 716) respondents were excluded from the dataset. As a result of outlier analysis performed on the dataset, 159 outliers were detected and excluded from the dataset to obtain healthier results from the analysis (n = 557). After excluding the extreme values, measures of central tendency plus skewness and kurtosis coefficients were calculated for the remaining dataset. Skewness and kurtosis values were ranged between -1.50 and +1.50.

Results of multicollinearity and singularity analyses indicate that the highest VIF (Variance Inflation Factors) value is 4.79 and smaller than 5, and the smallest Tolerance value is 0.21 and not smaller than 0.20 proving that no multicollinearity exists in the dataset (Alpar, 2014, p.415). Moreover, the Durbin-Watson value is 1.92l, which means that the values ranging between 1<Durbin-

Watson <3 show the absence of autocorrelation problem in the dataset and that the errors are independent of each other (Seçer, 2015; p. 150). At this point, the significance of the Kaiser-Meyer-Olkin (KMO) value and Barlett's Test of Sphericity were examined, where the KMO value was 0.92 and the Barlett's test proved significant (x^2 = 10215.35, *df* = 561, *p* <.05).

Principal component analysis

After ensuring the assumptions above, a principal component analysis was undertaken in the exploratory factor analysis. Here, the "principal components" analysis method was selected and the scale was limited to four factors in the context of functionality suggested in the theoretical framework (i.e. praise statements, negative feedback, effort-based feedback, and ability-based feedback). Since no relationship is assumed to exist between the constructs of the theoretically developed scale, of orthogonal rotation techniques, the "varimax" rotation was preferred. Preliminary analysis found 18 items with factor loadings of below 0.32, cross-loading on more than one factor with adequate factor loadings, and cross-loading on more than one factor with a difference not less than 0.20. These items were excluded and the factor analysis was repeated.

When looking at the "Communalities" table, the minimum value for each item in the "Extraction" column, which shows the variance explained by the items, must be 0.10. The smallest value in these variance values was 0.38. In addition, the scale was of a four-factor solution with an eigenvalue of minimum 1 accounting for 55.73% of the variance, and each of the sub-factors accounting for at least 5% of the variance. Accordingly, the eigenvalues of the factors and their explained variance are 4.894 and 14.39% for the first, 7.1 and 20.90% for the second, 4.189 and 12.32% for the third, 2.760 and 8.12% for the fourth factor, respectively, as they are shown in Table 3.

Table 3. Item factor loads of the scale, variances explained by subscales and item analysis

Items	1. Factor Praise Statements	2. Factor Negative Feedback	3. Factor Effort-Based Feedback	4. Factor Ability-Based Feedback	Item-Total Correlation
item 23	.716				.583
item 10	.703				.639
item 2	.697				.577
item 27	.677				.628
item 22	.676				.441
item 11	.649				.595
item 3	.643				.544
item 30	.635				.592
item 1	.577				.474
item 19		.810			.565
item 17		.803			.520
item 13		.774			.535
item 7		.768			.530
item 29		.761			.484
item 14		.752			.504
item 4		.745			.539
item 5		.741			.498
item 18		.733			.528
item 28		.723			.414
item 21		.652			.220
item 12		.622			.387
item 34			.720		.535
item 24			.706		.406
item 6			.684		.327
item 9			.674		.465
item 20			.644		.480
item 8			.583		.225
item 31			.582		.558
item 25			.528		.557

Table 3 Continued

Items	1. Factor Praise Statements	2. Factor Negative Feedback	3. Factor Effort-Based Feedback	4. Factor Ability-Based Feedback	Item-Total Correlation
item 15			.486		.306
item 32				.827	.415
item 33				.796	.425
item 26				.670	.471
item 16				.567	.486
	% 14.39	% 20.90	% 12.32	% 8.12	
	Total Variance: % 55.73				

Item-total correlation and factor loading for each item were 0.32 and above, and the items did not have an overlapping trait. Item-total correlation for praise statements sub-dimension ranged between 0.441 and 0.639, while the factor loadings ranged between 0.577 and 0.716. Item-total correlation for negative feedback sub-dimension ranged between 0.220 and 0.565, and the factor loadings between 0.622 and 0.810. Item-total correlation for effort-based feedback sub-dimension ranged between 0.225 and 0.558, and the factor loadings between 0.486 and 0.720. Lastly, the item-total correlation for ability-based feedback sub-dimension ranged between 0.415 and 0.486, and the factor loadings between 0.567 and 0.827. Figure 2 shows that a four-factor construct is provided since the values factors have after the fourth point are small as well as the distance between them is similar.

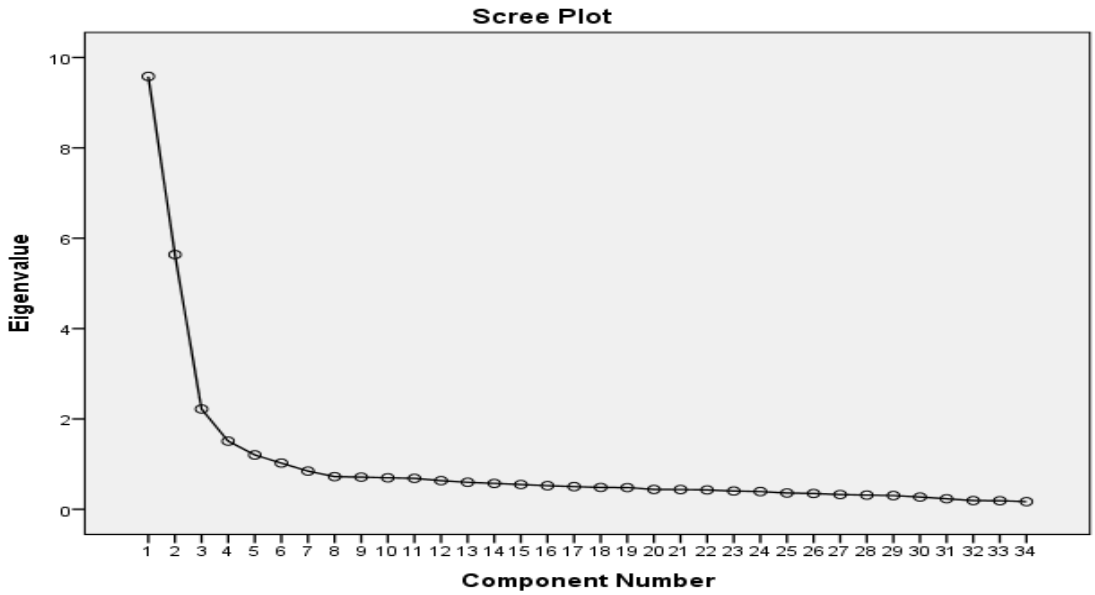

Figure 2. Scree plot

Confirmatory factor analysis (CFA)

Confirmatory factor analysis was conducted using relevant statistical package programs based on data from 761 students different from the group where the exploratory factor analysis was performed. To examine the model fit, the ratio of χ^2/df, RMSEA (Root Mean Square Error of Approximation), GFI (Goodness of Fit Index), CFI (Comparative Fit Index), and NNFI (Non-Normed Fit Index) values were examined. Goodness-of-fit statistics for the Science Course Feedback Perception Scale was found to be $\Delta\chi^2 = 1827.41$, $p = 0.00$, $df = 521$, $\chi^2/df = 3.50$, RMSEA = 0.05, GFI = 0.88, CFI = 0.93, NNFI = 0.92. It is suggested to use multiple fit indices to examine the model fit because the fit indices have weaknesses and strengths compared to one another (Kline, 2011). According to the goodness-of-fit criteria suggested by Schumacker and Lomax (2010), the ratio of

χ^2/df and other fit indices (CFI, GFI, and NNFI) are within the acceptable ranges. In addition, taking notice of appropriate modification suggestions, the ratio of χ^2/df fell below 3, and other fit indices improved. At the same time, in the diagram "for t values", it is stated that there is no red arrow; If there is a red arrow at the "t value", it is stated that the item in question means that it is not significant at 0.05 level and should be removed from the scale. However, the fact that no item was shown in red in the model showed that all items were compatible with the tested model. In addition, a path diagram for confirmatory factor analysis is given in Figure 3.

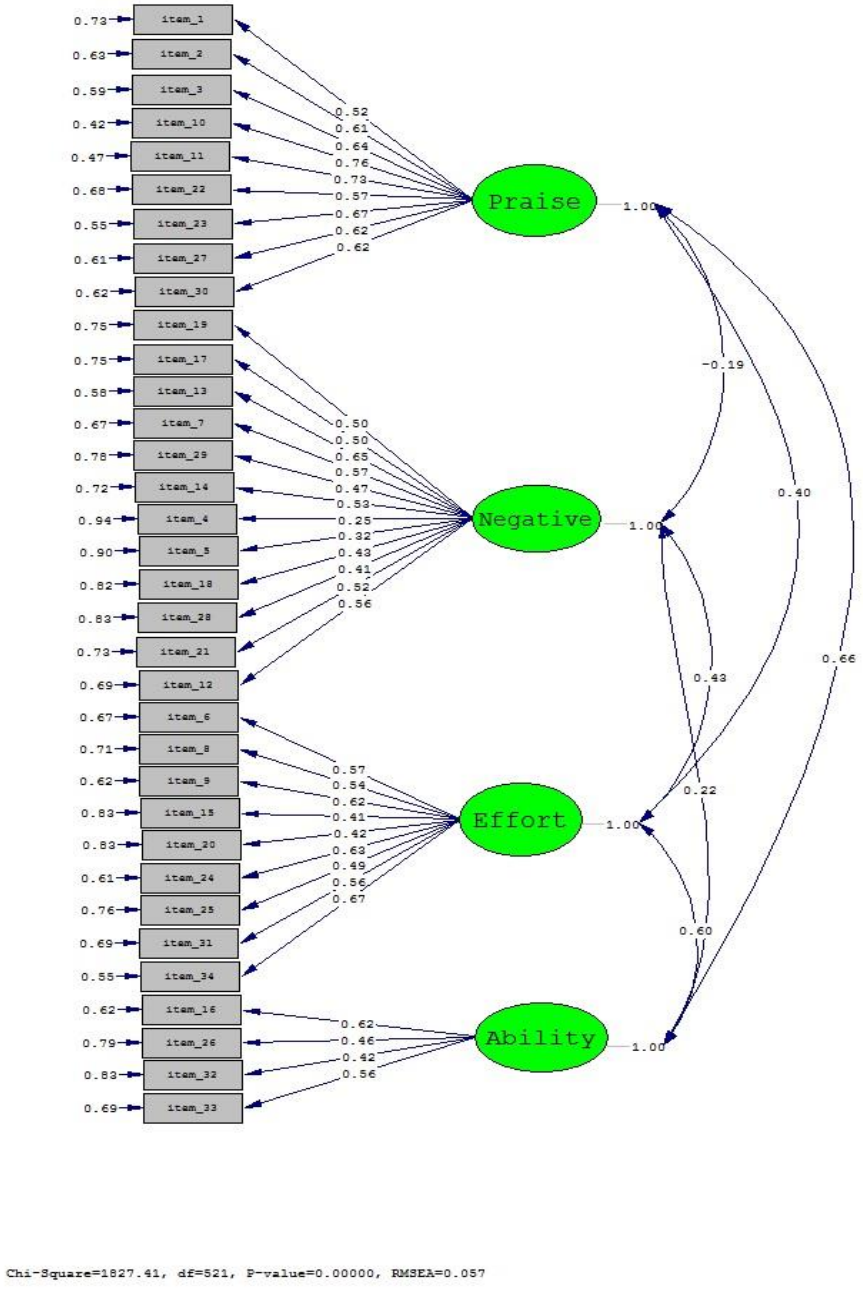

Figure 3. Path diagram

Validity and Reliability

Cronbach's Alpha reliability coefficient was calculated to determine the reliability of the scale so as to ensure how accurately the scale measures an intended feature (students' perception of feedback in this scale) in the quantitative part of the study. According to Büyüköztürk (2008), this coefficient is a measure of the consistency of the scores belonging to the item with the total test scores. When this coefficient is between 0.60 and 0.80, the scale is quite reliable, and when between 0.80 and 1.00, the scale is highly reliable (Akgül & Çevik, 2003, p. 436). Accordingly, the findings on the reliability analysis of the scale are given in Table 4.

Table 4. The reliability analysis findings of science course feedback perception scale

Factors	Cronbach's Alpha Reliability Coefficient	Split-Half Reliability
Praise Statements	0.884	0.862
Negative Feedback	0.927	0.813
Effort-Based Feedback	0.834	0.820
Ability-Based Feedback	0.823	0.722
Total Scale	0.913	0.850

Cronbach's Alpha reliability coefficient for the whole scale was calculated as $\alpha = 0.913$ in the analysis. Internal consistency coefficients of the scale were found as $\alpha=0.884$ for the first factor (praise statements), $\alpha=0.927$ for the second (negative feedback), $\alpha=0.834$ for the third (effort-based feedback), and $\alpha=0.823$ for the fourth (ability-based feedback). These results could be shown as evidence that the scale is reliable and that the internal consistency coefficients of the sub-dimensions of the scale are also quite high. Besides, the split-half reliability of the scale was calculated as $r = 0.850$. The split-half reliability was calculated as $r = 0.862$ for the first factor (praise statements), $r = 0.813$ for the second (negative feedback), $r = 0.820$ for the third (effort-based feedback), and $r = 0.722$ for the third (ability-based feedback). Considering that a scale with a reliability value of 0.70 is adequate in scale development and adaption process, it could be argued that the scale has met the split-half reliability criteria.

205 people who constitute the first 27% of the 761 group are designated as the upper group, while 205 people who make up the last 27% are the subgroup. The differential validity of the scale was examined by analyzing the four sub-factors in the scale based on the difference of the lower-upper slice group averages. While applying this method, the scale scores of the participants were firstly sorted from large to small. For each scale item of the 27% lower-upper groups at both ends of the distribution of scale points, the difference between the t test for the independent groups and their averages was examined. The t test results showed that the mean score of the upper 27% group was significantly higher than the same score of the lower 27% group ($p < .001$) in all items. The values obtained from this analysis are presented in Table 5.

Table 5. t values for 27% of lower and upper groups

Factor	Items	t-values	Factor	Items	t-values
Praise Statements	item1	21.398*	Effort-Based Feedback	item 6	19.038*
	item 2	23.091*		item 8	14.008*
	item 3	21.415*		item 9	22.129*
	item 10	30.952*		item 15	14.287*
	item 11	26.659*		item 20	19.240*
	item 22	24.266*		item 24	23.026*
	item 23	28.907*		item 25	15.393*
	item 27	23.617*		item 31	21.205*
	item 30	22.768*		item 34	23.281*
Negative Feedback	item 19	13.018*	Ability-Based Feedback	item 16	22.118*
	item 17	12.863*		item 26	17.175*
	item 13	18.701*		item 32	10.936*
	item 7	14.894*		item 33	14.314*

item 29	11.770*
item 14	14.915*
item 4	5.992*
item 5	8.199*
item 18	14.687*
item 28	9.640*
item 21	18.401*
item 12	19.903*

*$p < 0.01$

Findings from the quantitative data analysis

Findings related to the first sub-problem: "What are the feedback scores of students (praise score, effort-based feedback score, negative feedback score, and ability-based feedback score)?

The arithmetic means and standard deviations of the scores the secondary school students of grades 6, 7, 8 obtained from the sub-dimensions of science course feedback perception scale and the lowest or the highest score that could be obtained from these sub-dimensions are given in Table 6.

Table 6. Table of arithmetic means and standard deviations of scores from the scale according to the types of feedback (*M*: arithmetic mean, *SD*: standard deviation)

Feedback Score Type	*M*	*SD*	Obtainable Score	
			Minimum Score	Maximum Score
Praise Statements	21.34	6.252	9	36
Effort-Based Feedback	16.05	4.117	9	36
Negative Feedback	18.58	5.485	12	48
Ability-Based Feedback	5.54	1.895	4	16

As Table 6 indicates, the lowest score of 9 and the highest score of 36 may be obtained from the sub-dimension of praise statements. However, the arithmetic mean of the scores the students received from this sub-dimension was $M = 21.34$ with a standard deviation of $SD = 6.252$. This score corresponds to the "often" range of group value. That is to say, students may hear praise statements frequently in the science class. As such, the lowest score of 9 and the highest score of 36 is obtainable from the effort-based feedback sub-dimension, but the arithmetic mean of scores the students obtained from this sub-dimension was $M = 16.06$ with a standard deviation of $SD = 4.117$. This score corresponds to the "sometimes" range of group value. Hence, it could be argued that students only "sometimes" hear effort-based feedback statements in the science class. Moreover, the lowest of 12 and the highest score 48 are obtainable from the negative feedback sub-dimension, whereas the students received an arithmetic mean of $M = 18.58$ and a standard deviation of $SD = 5.485$ for this sub-dimension. This score corresponds to the "never" range of group value. That is to say, students rarely hear negative feedback as to say they never hear such feedback in the science class. Lastly, the lowest score of 4 and the highest score of 16 are obtainable from the sub-dimension of ability-based feedback, while the students received an arithmetic mean of $M = 5.54$ and a standard deviation of $SD = 1.895$ from this sub-dimension. This corresponds to the "never" range of group value. So, it could be argued that students seldom hear ability-based feedback in the science class. Considering the sub-dimensions of the science course feedback perception scale, the secondary school students of grades 6, 7, and 8 may often hear praise statements, sometimes hear effort-based feedback, and seldom hear negative and ability-based feedback statements.

Findings related to the Sub-Problem of: "Do the feedback scores of students significantly differ by gender?"

Given that all groups demonstrated non-normal distribution and the homogeneity of variance was not satisfied even if the normality was achieved, the Mann Whitney U test, which is the non-

parametric equivalent of the t-test, was applied in response to this sub-problem. The Mann Whitney U test results of scores that male and female students received from all feedback statements are given in Table 7.

Table 7. Mann Whitney U test result of scores obtained from all feedback statements in terms of gender

Sub-Dimensions	Gender	*n*	*Mean Rank*	*Total Rank*	U	*p*
Praise Statements	Female	437	372.72	162880.00	67177.00	.227
	Male	324	392.16	127061.00		
Effort-Based Feedback	Female	437	361.39	157926.50	62223.50	.004**
	Male	324	407.45	132014.50		
Negative Feedback	Female	437	342.31	149588.00	53885.00	.000**
	Male	324	433.19	140353.00		
Ability-Based Feedback	Female	437	359.84	157249.50	61546.50	.001**
	Male	324	409.54	132691.50		

*$p < 0.05$, **$p < 0.01$

According to Table 7, the Mann Whitney U test results yielded no significant difference between the scores of male and female students as regards praise statements (U = 67177.00, $p > 0.05$). However, the effort-based (U = 62223.50, $p < 0.05$), negative (U = 53885.00, $p < 0.05$), and ability-based feedback (U = 61546.50, $p < 0.05$) scores of students yielded a significant difference by gender. As shown in Table 6, male students received more effort-based, negative, and ability-based feedback than did female students according to the mean ranks of male and female students. However, male and female students together received praise statements almost equally.

Findings related to the sub-problem of: "Do the feedback scores of students significantly differ by grade level?"

Since the praise and effort-based feedback scores of students satisfied the normality assumption by subgroups (i.e. by grade levels), it was decided to conduct a one-way variance analysis to determine if these scores significantly differ by grade level. However, according to the equality of homogeneity test results (Levene test), the homogeneity assumption of ANOVA was violated for the praise scores ($p < 0.5$), but not for the effort-based feedback scores ($p > 0.5$). Therefore, since the praise scores were not homogeneous, Tamhane's T2 test was conducted to determine the mean scores from which the statistically significant difference arise (Can, 2014, p.149), and the Scheffe test was applied for the effort-based feedback scores, for they were homogeneous. The results of one-way variance analysis are given in Table 8.

Table 8. Means, standard deviations, and one-way ANOVA results of the scores obtained from the praise statements and effort-based feedback statements according to the grade levels

Sub-Dimensions	Grade Level	*n*	*M*	*SD*	*df*	*F*	*p*
Praise Statements	6. sınıf	276	20.70	5.82	2/758	7.877	.000**
	7. sınıf	246	22.63	5.98			
	8. sınıf	239	20.74	6.79			
Effort-Based Feedback	6. sınıf	276	17.93	5.41	2/758	3.107	.052
	7. sınıf	246	18.98	5.49			
	8. sınıf	239	18.93	5.50			

*$p < 0.05$, **$p < 0.01$

As seen in Table 8, the praise scores significantly differed by grade level ($F_{760} = 7.877$, $p<.05$). According to Tamhane's Test results, performed to trace the source of this difference in the mean scores, arithmetic mean of praise scores of 7th graders ($M = 22.63$) was significantly higher than that of 6th ($M = 20.70$) and 8th ($M = 20.74$) graders ($F_{760} = 7.877$ $p<0.05$). Accordingly, considering the arithmetic mean of praise scores, 7th graders received more praise statements than did the 6th and 8th graders. Further, there was no significant difference between the 6th and 8th graders in terms of teachers' use of praise statements. As shown in Table 7, the effort-based feedback scores of students did not yield a significant difference by grade level ($F_{760} = 3.107$, $p > 0.05$). This result suggests that students receive effort-based feedback at a relatively equal level. Since the test was not significant, the Scheffe test was not performed.

As negative and ability-based feedback scores of students did not meet the normality assumption, the Kruskal Wallis H test was conducted to determine if these scores significantly differ by grade level. The test results are given in Table 9.

Table 9. Kruskal Wallis H test result of scores obtained from the negative feedback and ability-based feedback statements according to grade levels

Sub-Dimensions	Grade Level	*n*	Mean Rank	*df*	χ^2	*p*
Negative Feedback	6. sınıf	276	391.07	2	1.708	.426
	7. sınıf	246	366.63			
	8. sınıf	239	384.15			
Ability-Based Feedback	6. sınıf	276	378.68	2	14.220	.001**
	7. sınıf	246	417.56			
	8. sınıf	239	346.04			

*$p < 0.05$, **$p < 0.01$

As seen in Table 9, the negative feedback scores of the students did not yield a significant difference by grade level ($\chi^2 = 1.708$, $p > 0.05$). Accordingly, the feedback teachers give to the students is at a relatively same level. By contrast, the ability-based feedback scores of students yielded a significant difference by grade level ($\chi^2 = 14.220$, $p < 0.05$). To determine the source of this difference in the mean ranks, the Mann Whitney U test was conducted. As a result of the test, a significant difference was found between the 8th and 7th grades ($p < 0.05$), whereas no significant difference was found between grades 6 and 8 or grades 6 and 7 ($p > 0.05$). According to these results, 7th graders received more ability-based feedback than did the 8th graders, while the 6th graders received almost the same amount of ability-based feedback as did the 7th and 8th graders.

DISCUSSION AND CONCLUSION

The study was conducted using exploratory sequential design of mixed-methods research to determine the feedback strategies used by science teachers in the classroom setting. Meanwhile, the exploratory sequential design is also known as scale development design. For this reason, feedback statements obtained in the qualitative dimension were used to develop the science course feedback perception scale. As a result of the study, the scale consisted of a total of 34 items including 9 praise statements, 9 effort-based feedback statements, 12 negative feedback statements, and 4 ability-based feedback statements. This way, the scale indicated a four-factor construct that accounted for 55.73% of the variance. Besides, Cronbach's Alpha reliability coefficient for the whole scale was $\alpha = 0.913$, $\alpha = 0.884$ for the praise statements sub-dimensions, $\alpha = 0.927$ for the negative feedback sub-dimension, $\alpha = 834$ for the effort-based feedback sub-dimension, and $\alpha = 0.823$ for the ability-based feedback sub-dimension. The items in the scale were organized in 4-point Likert type design with response categories of "Always, "Often", "Sometimes", and "Never". However, Burnett's (2002) Teacher Feedback Scale consists of 18 items including 5 items of praise statements, 5 items of negative

feedback, 4 items of effort-based feedback, and 4 items of ability-based feedback showing a four-factor construct. The scale is prepared in the form of a Likert type rating scale with three response categories of "Often", "Sometimes", and "Never". The amount of variance explained is not reported but the Cronbach's Alpha value for the praise statements sub-factor is $\alpha = 0.85$, $\alpha = 0.77$ for the negative feedback sub-factor, $\alpha = 0.78$ for the effort-based feedback sub-factor, and $\alpha = 0.79$ for the ability-based feedback sub-factor. When the two scales are compared, Cronbach's Alpha internal consistency coefficients for the praise statements sub-dimension are almost the same in both scales. By contrast, the scale developed in this study seems to have higher values in effort-based feedback, negative feedback, and ability-based feedback sub-dimensions. In addition, the scale Burnett (2002) developed was of a 3-point Likert scale, while the scale developed in this study was prepared in 4-point Likert type design. Again, Burnett's (2002) scale consists of 18 items and four factors, but the scale developed consists of 34 items and four factors. Followed by exploratory factor analysis, confirmatory factor analysis was performed based on responses given by 761 students of grades 6, 7, and 8. Since the fit indices were adequate, analyses were performed regarding the quantitative questions. Feedback statements obtained in the qualitative part of the study were used to develop the scale and statistical descriptive analyses were conducted using the scale developed. Considering the data obtained from the science course feedback perception scale, the secondary school students of grades 6, 7, and 8 were found to often hear statements of praise in their science class. Hitz and Driscoll (1994) considered praise to be a constructive reinforcement, believing that consistent praise encourages desirable behavior but eliminates the undesirable one. While praise statements are good motivators for the students, they are not enough when the definition and function of feedback are carefully taken into account. This is because Hattie and Timperley (2007) outline the questions a feedback statement should answer are: *(i)* Where am I going? *(ii)* How am I doing? *(iii)* What is my next step? Therefore, feedback statements based on appreciating student behaviors alone will not be able to provide adequate answers to the questions asked. Also, students were found to sometimes hear effort-based feedback statements. The kind of teaching that aligns with a good learning process is known to be a process-oriented education, which involves guiding and facilitating student learning (Vermunt, 1992). Statements that address the facilitating aspect of learning are the statements that encourage students, and these statements are often effort-based feedback statements. Hitz and Driscoll (1994) argue that constructive encouragement is needed to increase self-confidence, and this should focus on betterments and efforts using sincere comments and recognizing students' feelings. Using this type of feedback more in the classroom setting will be positive for students.

Another finding in this study is that students rarely receive negative and ability-based feedback. A low level of negative feedback is important for students' motivation, as negative feedback leads to negative feelings and lowers their performance (London, 1995). Feedback on affective learning activities includes comments focusing on encouragement, effort, attention, and dealing with emotions, in that feedback is also known to affect affective learning activities. Affective feedback can have an indirect effect on what students learn only via its influence on cognitive learning activities (Vermunt & Verloop, 1999). Affective feedback is always a part of interpersonal communication between the teacher and student (Meyer & Turner, 2002). These and similar statements can damage communication between the teacher and the student. However, ability-based feedback statements are significant feedback types for students, which can increase both the self-confidence and self-efficacy of the student. To give such feedback type, teachers are required to know their students closely. Therefore, one of the reasons why ability-based kind of feedback is rarely given may stem from teachers' inadequate awareness of students' abilities. Students praise statement scores did not significantly differ by gender. However, effort-based feedback, negative feedback, and ability-based feedback scores of male students were higher than that of females. Merrett and Wheldall (1987) observed 128 teachers in England and found that 56% of the feedback given to students was positive, while 44% was negative. Additionally, males were more likely to receive negative responses from female teachers for social behaviors, while male teachers gave positive responses to male students for academic behavior. The reason for the absence of gender difference in praise statements may be that praise statements do not provide enough guidance to the student in terms of content. Because it has already been noted that feedback has two dimensions of content and timing and that the content

dimension is particular information for the student showing how far s/he is from the expected performance to achieve (Van de Ridder et al., 2008). As statements of praise were "Yeah", "OK", "Congratulations" and so forth, they may not have provided much guidance for students of different levels in terms of factors such as readiness and motivation besides gender. However, praise statement scores of 7th graders were higher than that of the 6th and 8th graders, that is, the 7th graders were found to hear praise statements in the classroom setting more frequently than did the 6th and 8th graders. There was no statistically significant difference between the effort-based feedback scores of students according to their grade levels. Similarly, there was no statistically significant difference between the negative feedback scores of students according to their grade level.

When the ability-based feedback scores of students are examined by grade level, the ability-based feedback scores of 7th graders were higher than that of the 8th graders, meaning that the 7th graders seem to hear ability-based feedback more often than do the 8th graders. Burnett (2002), in his study, found that as students' grade levels increased, the number of those seeking ability-based feedback decreased, while the number of those seeking effort-based feedback increased. Besides, in their study, Burnett and Mandel (2010) found that teachers recommend using ability-based feedback rather than effort-based feedback statements in younger age groups (grade 1-4), while emphasized using effort-based feedback rather than ability-based feedback statements in older age groups (grade 5-7).

Recommendations

Considering recommendations that could be given for classroom applications, science teachers were found to most frequently use praise statements in the classroom setting. However, they should focus more on using effort-based feedback statements in the classroom. Following a constructivist approach, teachers need to prefer feedback statements that can facilitate students' learning and transfer them to new learning situations. Professional development programs could be organized to raise the awareness of science teachers about feedback. Therefore, the scale developed in this research could be used for conducting a needs analysis. It was discovered that science teachers give more negative, effort-based, and ability-based feedback to male students than to females. Teachers are therefore recommended to use less negative feedback and protect gender balance when giving effort-based feedback. Future studies could use a qualitative approach to investigate *(i)* why students barely receive ability-based feedback and why the 7th graders receive more praise-based feedback. *(ii)* The relationship between the students' perceived feedback in science course and their academic achievement, attitude, and scientific process skills could be investigated.

REFERENCES

Adrienne, R. (1997). Feedback: Enhancing the performance of adult learners with learning disabilities, *National Adult Literacy and Learning Disabilities Center*, Washington, DC.

Akgül A. ve Çevik O. (2003). *İstatistiksel Analiz Teknikleri [Statistical Analysis Techniques]*. Emek Ofset: Ankara

Alpar, R. (2014). *Uygulamalı istatistik ve geçerlik-güvenirlik [Applied Statistics and Validity-Reliability]*. Ankara: Detay Yayıncılık.

Blote, A.W. (1995). Students' self-concept in relation to perceived differential teacher treatment. *Learning and Instruction,* 5, 221–236.

Brinko, K. T. (1990). *Optimal conditions for effective feedback*, Paper Presented to the Annual Meeting of the American Educational Research Association, Boston, 16-20 April.

Burnett, P.C. (1996). Children's' self-talk and significant others' positive and negative statements. *Educational Psychology*, 16, 57-68.

Burnett, P.C. (2002). Teacher praise and feedback and students' perceptions of the classroom environment. *Educational Psychology: An International Journal of Experimental Educational Psychology, 22*(1), 5-16. https://doi.org/10.1080/01443410120101215

Burnett, P.C.,& Mandel, V. (2010). Praise and feedback in the primary classroom: teachers' and students' perspectives. *Australian Journal of Educational & Developmental Psychology, 10*, 145-154.

Büyüköztürk, Ş. (2008). *Sosyal bilimler için veri analizi el kitabı [A data analysis handbook for social science]*. Ankara: Pegem A Yayıncılık.

Büyüköztürk, Ş. (2015). *Sosyal bilimler için veri analizi el kitabı [A data analysis handbook for social science]*. Ankara: Pegem A Yayıncılık.

Can, A. (2014). *Spss ile bilimsel araştırma sürecinde nicel veri analizi [Quantitative data analysis in scientific research with SPSS]*. Ankara: Pegem A Yayıncılık.

Cengiz, E. (2015). *Fen bilimleri dersinde öğrenci hataları ve öğretmenlerin bu hatalara verdikleri geri bildirimlerin incelenmesi [Investigating Mistakes Students Make and the Feedback Teachers Provide on these Mistakes]*. Yüksek lisans tezi. Karadeniz Teknik Üniversitesi Eğitim Bilimleri Enstitüsü,Trabzon.

Comrey, A.L.,& Lee, H.B. (1992). A first course in factor analysis. *Multivariate Behavioral Research, 1*, 245-276.

Creswell, J.W.,Fetters, M.D., & Ivankova, N.V. (2004) Designing a mixed methods study in primary care. *The Annals of Family Medicine, 2*, 7-12.

Çıngı, H. (1994). *Örnekleme kuramı [Theory of Sampling]*. H.Ü. Fen Fakültesi Yayınları, Ankara.

Çobakçor, B.Ö.,Akşan, E., Öztürk, T. ve Çimer, S.O. (2011). İlköğretim matematik öğretmeni adaylarının matematik derslerinden aldığı ve tercih ettiği geri bildirim türleri [Types of Feedback Prospective Mathematics Teachers Receive and Prefer in Mathematics Course]. *Turkish Journal of Computer and Mathematics Education, 2*(1), 46-68.

Dökmen, Ü. (1982). Farklı tür geri bildirimlerin öğrenmeye etkisi [The Effect of Different Types of Feedback on Learning]. *Ankara Üniversitesi Eğitim Bilimleri Dergisi*, 71-80.

Field, A. (2000). *Discovering statistics using SPSS: Advanced techniques for beginners*. London. Sage Publication.

Hattie, J. (2003).Teachers make a difference: What is there search evidence? Paper presented at the Building Teacher Quality: What does the research tell us ACER Research Conference, Melbourne, Australia. Retrieved from http://research.acer.edu.au/research_conference_2003/4/

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112. https://doi.org/10.3102/003465430298487

Hattie, J. (2009). V*isible learning: A synthesis of over 800 meta-analyses relating to achievement*. London and New York: Routledge.

Higgins, R. (2000)."*Be More Critical!" Rethinking assessment feedback*. Paper Presented at The BERA Conference Cardiff University, September 7-10.

Hitz, R. & Driscoll, A. (1989). Praise or encouragement? New insights into praise. Implications for early childhood teachers. *Young Children, 43,* 6–13.

Hutcheson, R. K. & Sofroniou, N. (1999). *The multivariate social scientist*. London: Sage Publications.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. NY: Guilford Publication, Inc.

London, M. (1995). Giving feedback: Source-centered antecedents and consequences of constructive and destructive feedback. *Human Resource Management Review, 5,* 159–188.

Lynn, M.R. (1986). Determination and quantification of content validity. *Nursing Research*, 35, 382–385.

Köğce, D. (2012). *İlköğretim matematik öğretmenlerinin geri bildirim verme biçimlerinin incelenmesi [Investigating primary mathematics teachers' ways providing feedback]*. Doctoral dissertation. Karadeniz Teknik University, Institute of Educational Science, Trabzon.

Meyer, D. K.,&Turner, J. C. (2002). Discovering emotion in classroom motivation research. *Educational Psychologist, 37,* 107–114. https://doi.org/10.1207/s15326985ep3702_5

Merrett, F. & Wheldall, K. (1987).Natural rates of teacher approval and disapproval in British primary and middle school classrooms. *Educational Psychology. 57*(1), 95-103. https://doi.org/10.1111/j.2044-8279.1987.tb03064.x

Miles, M. B. & Huberman, A.M.(1994). *Qualitative data analysis*.(2nd).Thousand Oak: SAGE

Miles, M.B. & Huberman, A.M. (2015). *Qualitative data analysis* (S. Akbaba Altun & A. Ersoy, Çev.). (2nd). Pegem Akademi. Ankara

Morgan, G. A, Leech, N. L., Gloeckner, G. W., & Barret, K.C. (2004). *SPSS for Introductory Statistic: Use and Interpretation*. (2nd Edition). London: Lawrence Erlbaum Associates.

Orlich, C. O., Harder, R. J., Callahan, R. C., Trevisan, M. S., & Brown, A. H. (2007). *Teaching strategies: A guide to effective instruction (8th)*. New York: Houghton Mifflin Company

Pallant, J. (2001). *SPSS survival manual*. Maidenhead, PA: Open University Press.

Patton, M.Q. (2002). *Qualitative research and evaluation methods*. (3rd). Thousand Oak: SAGE

Sadler, D. R. (1989). Formative assessment and the design of instructional systems, *Instructional Science*. *18*(2), 119-144.

Sauls, D. J. (2004). The Labor Support Questionnaire: Development and psychometric analysis. *Journal of Nursing Measurement*, 12, 123–312. https://doi.org/10.1891/jnum.2004.12.2.123

Schumacker, R. E. & Lomax R. G.(2010). Beginner's guide to structural equation modeling (3rd edition). NY: Routledge

Seçer, İ. (2015). *Spss ve Lisrel ile pratik veri analizi: Analiz ve raporlaştırma [Practical Data Analysis with SPSS and Lisrel: Analyzing and Reporting]*. Ankara: Anı Yayıncılık.

Shepell, W. (2000). Health Quest: A quarterly news letter focusing on mental health issues and concerns. Retrieved from https://www.utdallas.edu on 03.05.2017

Strauss, A. L. & Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, CA: Sage

Tabachnick, B. G. & Fidell, L. S. (1996). *Using multivariate statistics* (3rd Ed.). New York: Harper Collins.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5th Ed.). New York: Allynand Bacon.

Tabachnick, B.G. & Fidell. L. S. (2013). *Using Multivariate Statistics* (Sixth Ed.) Pearson, Boston.

Taras, M. (2005). Assessment summative and formative some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466–478.

Tashakkori A. & Cresswell J. W. (2007). The new era of mixed methods. *Journal of Mixed Methods*, 1, 2-8.

Thalheimer, W. & Cook, S., (2002). How to calculate effect sizes from published research: A simplified methodology. *Work-Learning Research*, 1-9.

Türkdoğan A. ve Baki A. (2012). İlköğretim ikinci kademe matematik öğretmenlerinin yanlışlara dönüt vermede kullandıkları dönüt teknikleri [Feedback techniques used by middle school mathematics teachers to give feedback to mistakes]. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, *45*(2), 157-182.

Van De Ridder, J. M. M., Stokking, K. M., McGaghie, W. C., & Ten Cate, O. T. J. (2008). What is feedback in clinical education? *Medical education*, *42*(2), 189-197. https://doi.org/10.1111/j.1365-2923.2007.02973.x

Vermunt, J. D. & N. Verloop. (1999). Congruence and friction between learning and teaching. *Learning and Instruction, 9*, 257–280.

Vollmeyer, R. & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction*, *15*(6), 589-602. https://doi.org/10.1016/j.learninstruc.2005.08.001