International Journal of Assessment Tools in Education
2026, Vol. 13, No. 2, 587–602
https://doi.org/10.21449/ijate.1683080
journal homepage: https://dergipark.org.tr/en/pub/ijate *Research Article*

Developing a teacher collaboration scale based on item response theory

Yeliz Özkan Hıdıroğlu [1*], **Eren Can Aybek** [2]

[1]Pamukkale University, Faculty of Education, Department of Science and Mathematics Education, Denizli, Turkiye
[2]Pamukkale University, Faculty of Education, Department of Measurement and Evaluation, Denizli, Turkiye

ARTICLE HISTORY

Received: Apr. 24, 2025

Accepted: Feb. 14, 2026

KEYWORDS

Item response theory,
Graded response model,
Scale development,
Teacher collaboration.

Abstract: The purpose of this study is to develop a psychometrically sound Teacher Collaboration Scale based on Item Response Theory. The data obtained from 461 teachers were analyzed, and items were calibrated using Samejima's Graded Response Model. Exploratory Factor Analysis confirmed a unidimensional structure explaining 53.51% of the variance, while local independence and item-model fit were rigorously tested. The final 18-item scale demonstrated high discrimination parameters ranging from 1.815 to 4.777. The analysis of the test information function indicated that the scale provided the highest precision and reliability for teachers with collaboration levels between -3 and +1 theta. Additionally, Differential Item Functioning analysis revealed no gender bias, confirming the scale was a valid and reliable instrument.

INTRODUCTION

Teachers have great responsibilities in preserving civilization, developing and transferring culture to younger generations, strengthening the beliefs and value judgments of the society, gaining development and development consciousness as a social consciousness and establishing a democratic social order. Teachers are the most important human elements of education that strongly determine learning success in schools. They are the ones who enable both learning, utilizing knowledge and obtaining knowledge from knowledge. They play an important role in the development of society by influencing the perspectives of young people (Ahrari *et al*., 2021).

Since the technological and scientific developments in each period are different, the teacher profiles, competencies and skills needed in that period differ. In order to equip future students with 21st-century skills, the competencies of teachers who will provide them with these skills are an important research topic. Collaboration is also an important skill in addition to 21st-century skills such as critical thinking, communication, creativity, various literacies (digital literacy, technology literacy, financial literacy, media literacy, academic literacy, etc.), computational thinking, metacognitive thinking, and interdisciplinary thinking. Collaboration is included among learning and innovation skills in The Partnership for 21st Century Learning Frameworks (2009) and among interpersonal skills in The National Research Council (2012).

***Corresponding author at:** Yeliz ÖZKAN HIDIROĞLU yelizozkan09@gmail.com Pamukkale University, Faculty of Education, Department of Science and Mathematics Education, 20070, Denizli, Turkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

e-ISSN: 2148-7456

Statement of Problem

Co-operation theory

Chester Barnard is the founder of the Co-operation theory. In his book The Functions of Executive, he described the organization as a physiological, social, biological and psychological system with systematic relationships between them and defined it as a system of cooperation in which two or more individuals consciously coordinate activities (Barnard, 1958). He emphasizes the importance of informal organizations and examines the organization as a system. Barnard (1958) claims that the effective and efficient functioning of the organization can only occur as a result of the effective cooperation of employees with each other and that the concept of voluntarism is an indispensable element for cooperation systems. There are three basic elements in the emergence of organizations: common purpose, communication and willingness of employees to cooperate (Bursalıoğlu, 2014). Cooperation can be realized when employees communicate with each other. According to Barnard, organizations can survive if they can motivate their employees and ensure employee cooperation (Wren & Bedeian, 2009).

Teacher collaboration

Teacher collaboration is vital for transforming schools into learning organizations and preparing students for the future. In business life, it is one of the most important skills for effective job performance (American Management Association, 2019; OECD, 2017). Teacher collaboration has not been recognized as an end in itself; it has been presented as a means for change, a prerequisite for improvement, and a way to discuss school and educational issues (Hooningh & Hooge, 2014). Therefore, it is important to clarify the definition of teacher collaboration and determine whether it is a means to achieve work-related outcomes or a vital skill to be acquired. Collaboration is developing and maintaining a common understanding of a problem through coordinated and simultaneous activities (Roschelle & Teasley, 1995). Participants in organizational actions share the same goal (Bursalıoğlu, 2000). In a similar definition, collaboration is a set of activities in which a common understanding of the goal is coordinated, a flexible responsibility and role sharing are realized. Effective collaboration between teachers is important in transforming students into competent collaborators of the future, as it serves as a role model for students' collaborative work (Coke, 2005). What is achieved through teacher collaboration is more than what is achieved as lone workers. In the "General Qualifications Document for the Teaching Profession" published by the Ministry of National Education (MoNE), it is stated that it is important for teachers to collaborate with their colleagues and collaboration is included as a competency area (MoNE, 2017). School principals also have important responsibilities in ensuring teacher collaboration (Meyer *et al.*, 2020). In a longitudinal study conducted with nine teachers in the United States, it was determined that it was not enough for principals to provide appropriate physical conditions for collaboration, but that principals "need to create a culture of shared vision, goals, and responsibility" (Szczesiul & Huizenga, 2014).

In high-performing countries such as Finland, teachers collaborate to a great extent and enjoy the benefits of collaboration. The benefits of teacher collaboration are identified in the literature at three levels. These are student level, teacher level and school level. At the student level, the benefits are improved student understanding and performance (Egodawatte *et al.*, 2011; Goddard *et al.*, 2007; Wigglesworth, 2011, Wullschleger *et al.*, 2025). At the teacher level, the benefits are increased morale and motivation, increased communication skills and technological equipment, and reduced workload (Bertrand *et al.*, 2006; Egodawatte *et al.*, 2011, Richter *et al.*, 2022), while at the organizational level, the benefits are renewal of the school climate, a harmonious and fair cultural change, school-wide attention to student needs, and the creation of a culture of intellectual inquiry (Euwema & Van der Waals, 2007; Khasawneh *et al.*, 2023; Moolenaar, 2010; Westheimer, 2008). In this direction, determining the subjects in which teachers cooperate and revealing their level of cooperation is an important research topic.

Teacher collaboration is critically important, as it contributes to teachers' professional development and supports the functioning and improvement of the school. Therefore, measuring teacher collaboration is essential for enhancing instructional quality and understanding the relationship between professional development and school structure (Özkan Hıdıroğlu & Tanrıöğen, 2020; Özkan Hıdıroğlu & Tanrıöğen, 2021). Reliable measurement facilitates comparisons across schools and is valuable for informing policymakers and school administrators. It also provides researchers with the necessary data to examine the effects of teacher collaboration.

Background and measures for teacher collaboration

When the literature is examined, it is seen that teacher collaboration is handled in different ways. Tschannen-Moran (1998) examined collaboration in schools and discussed collaboration under the titles of "collaboration between principals and teachers in school decisions", "collaboration with parents in school decisions," and "collaboration between teachers in classroom decisions." In his "Collaborative School Culture Survey," Gruenert (1998) examined collaboration with the dimensions of "unity of purpose", "collaborative leadership", "professional development", "teacher collaboration", "collegial support" and "learning partnership". Cerit (2009) examined collaboration as collaboration at school and discussed collaboration as "principal and teachers", "teachers and colleagues" and "school's relationship with parents". Goddard *et al.* (2010) developed a scale of teacher collaboration and identified three dimensions: "formal collaboration", "frequency of collaboration in teaching", and "teachers' collaboration on teaching policies". Dumay *et al.* (2013) addressed collaboration with the dimensions of "participation in school professional networks", "intellectual stimulation", "common vision", and "quality of professional relationships" in their teacher collaboration scale. Child and Shaw (2016) put forward a six-dimensional collaboration framework, including "mutual commitment", "conflict management", "introducing new ideas", "sharing resources", "task sharing", and "communication". Lench *et al.* (2015) presented a comprehensive framework with five dimensions: "self-awareness," "communication," "negotiation and decision-making," "contribution and support," and "observation and adaptation."

Çelebi *et al.* (2016) examined only the collaboration among the teachers belonging to the same department and analyzed the collaboration at three levels: "group formation", "early development and rulemaking", and "teaming". Yılmaz and Çelik (2020) developed a scale to reveal teachers' attitudes towards professional collaboration. Özüdoğru (2021) examined collaboration from the perspectives of "cooperation and need for cooperation", "obstacles to cooperation" and "suggestions for improving cooperation". Saylık and Arastaman (2022) examined collaboration from the perspective of collaboration at school and examined it with the dimensions of "school principal-stakeholder", "teacher-colleague" and "school guidance service-stakeholder" collaboration. Although previous studies have presented various classifications of teacher collaboration, none of these investigations have been examined within the framework of Item Response Theory.

Item response theory and the models used

In previous studies on teacher collaboration, measurement instruments were predominantly developed under the framework of Classical Test Theory (CTT). Although this preference is reasonable given the practical simplicity of CTT, these scales inherit several methodological limitations that restrict the precision and generalizability of the findings. Most notably, item statistics in CTT are sample-dependent, and a single reliability/standard error value is assumed for the entire test, making it impossible to evaluate measurement precision across different levels of the construct. Considering that teacher collaboration research increasingly involves comparisons across demographic groups and applications in diverse educational contexts, these limitations create challenges for producing stable and equitable scores.

For the present study, Item Response Theory (IRT) provides a stronger methodological foundation because it allows item parameters (difficulty and discrimination) and person parameters (θ) to be estimated independently, ensuring that the developed scale remains invariant across populations. This is particularly critical for teacher collaboration, a construct known to vary systematically across school types, experience levels, and institutional contexts. Moreover, IRT enables the examination of differential item functioning (DIF), test equating, and the potential future use of Computerized Adaptive Testing (CAT), all of which align with our long-term aim of developing a flexible and generalizable measurement instrument.

Among polytomous IRT models, the Samejima's (1996) Graded Response Model (GRM) was selected because it is theoretically consistent with the nature of Likert-type items used in this scale and is specifically designed to model ordered response categories with varying endorsement thresholds.

Purpose of the Study

The aim of this study is to develop the "Teacher Collaboration Scale", which was prepared by the researchers directly for teachers, based on IRT and to conduct a validity and reliability study. IRT was preferred because it offers item-level discrimination and difficulty parameters and produces more precise and stable estimates across different levels of the latent trait than CTT. The scales for teacher collaboration and the dimensions that emerged in these scales are given in detail, respectively (Çelebi *et al.*, 2016; Cerit, 2009; Child & Shaw, 2016; Dumay *et al.*, 2013; Goddard *et al.*, 2010; Gruenert, 1998; Lench *et al.*, 2015; Özüdoğru, 2021; Saylık & Arastaman, 2022; Tschannen-Moran, 1998; Yılmaz & Çelik, 2020). In this study, it was seen that there was no original scale developed by the researchers directly for teachers, addressing teacher collaboration holistically and developed in the local culture. For this purpose, the teacher collaboration scale was put forward, and the validity and reliability studies carried out in the process were stated in detail. The presence of validity gaps and psychometric weaknesses in the existing literature, as well as the use of outdated conceptual frameworks, makes this study particularly significant. This scale will provide an idea in determining the level of cooperation of teachers, who are one of the important stakeholders of school culture, revealing the current situation and taking measures for possible improvements. In addition, the relationship between teacher collaboration and various variables can be examined, which factors affect it more and which factors affect it more can be revealed. Since it is assumed that this study will meet the needs of experts, practitioners, researchers and academicians who conduct research on teacher collaboration, it is thought to contribute to the literature. It is thought that this scale will provide important data on collaboration and contribute to the literature as a unique scale for teacher collaboration.

METHOD

This study is a scale development study from quantitative research. Information about the sampling group, measurement tool and the techniques used in data analysis in the research to be carried out in the survey model is given below.

Research Group

In scaling studies conducted within the scope of item response theory, de Ayala (2009) noted that sample sizes of a few hundred participants are generally sufficient for testing the underlying assumptions. In this study, data were collected from 529 volunteer teachers working in Denizli, Turkiye. After removing multivariate outliers, the analyses were carried out with the remaining 461 teachers. Of these participants, 293 were female and 168 were male. The mean age of the teachers was 43.35, and the median age was 42.5. Regarding teaching levels, the sample included 31 preschool teachers, 101 primary school teachers, 196 middle school teachers, and 133 high school teachers.

Data Collection

The data were collected through the online administration of the Teacher Collaboration Scale. In the initial phase, a literature review was conducted, leading to the development of a 47-item draft form, and feedback was obtained from five experts. Four of these experts specialized in educational sciences (two full professors and two associate professors), while one expert specialized in mathematics education (one associate professor). Based on the expert evaluations, four items were removed from the form. Subsequently, the remaining items were reviewed by four language experts (Turkish language teachers) in terms of clarity and grammatical accuracy. Revisions were made to 16 items. Following these revisions, the 43-item scale was administered to the participants. Completion of the scale took approximately 10 minutes. All items were rated on a five-point Likert scale ranging from “Strongly Disagree (1)” to “Strongly Agree (5).” Additionally, demographic questions such as gender and age were included to describe the characteristics of the sample. The final form of the Teacher Collaboration Scale and all scale items are presented in Appendices.

Data Analysis

The data analysis is handled with three different phases. First, we prepared the data for analysis. Second, we checked the assumptions of IRT, and consequently, we calibrated the items with GRM. In the data preparation phase, we found that there were no missing values. So, we checked the multivariate outliers by comparing Mahalanobis distances with a Chi-square distribution with 43 degrees of freedom and accepted the alpha level as .001. According to the results, the data of 68 examinees were flagged as multivariate outliers and removed from the dataset. The data from the remaining 461 participants were checked for multivariate normality with the Henze-Zirkler test, and it was found that the data did not meet the multivariate normality assumption (HZ = 1.72; p < .05). The dp2fa software (Aybek, 2021) was used to detect and clean multivariate outliers and to check multivariate normality.

After the data were prepared for analysis, IRT assumptions were tested. For this purpose, unidimensionality was tested using exploratory factor analysis *(*EFA*)* (de Ayala, 2009). Since multivariate normality was not met, the principal axis factoring method was used as a factor extraction technique (Fabrigar *et al.,* 1999). Factor loading criterion was accepted as .40 and communality criterion as .30. While deciding the number of factors, parralel analysis, eigenvalues, and scree plot were investigated as combined. In addition, Kaiser-Meyer-Olkin value was used to determine the suitability of the sample size, and Bartlett's test of sphericity was used to see if factorizability is possible. Factor analysis was conducted with Jamovi (The Jamovi Project, 2022) software. Because we tested a unidimensional model, we did not use any axis rotation method, and the data of 461 participants were used while conducting EFA. According to the results of the exploratory factor analysis, five items were removed from the scale because they did not meet the factor loading and communality criteria.

Item calibrations for the remaining items were conducted using the Graded Response Model (Samejima, 1996). For this purpose, the mirt (Chalmers, 2012) package in R (R Core Team, 2022) was utilized. In addition, item model fits were determined by examining S_χ^2 values, and the assumption of local independence was checked by using Yen's Q3 statistic. Alpha level selected as .05 for interpreting the S_χ^2 while .37 was used as the cutoff value for the Q3 statistics (Boeschen Hospers, *et. al.*, 2016). After item calibrations, Option Response Functions, Item Information Functions, Test Information Function, Test Standard Error Function, and Marginal Reliability Function were examined. Finally, Differential Item Functioning was applied to determine the bias of the scale items.

The initial item pool was deliberately constructed to be substantially larger - approximately three times the targeted number of items - to enable item elimination and to identify the most valid and reliable items. Consequently, the removal of items did not compromise the conceptual coverage of the scale. The subsequent evaluation of the retained items demonstrated that they sufficiently represented the intended construct and were well aligned with the defined purpose of the scale.

FINDINGS

Exploratory Factor Analysis and Unidimensionality

For the EFA conducted to determine whether the 43 items in the draft scale were included under a single dimension, the KMO value was found to be .97 and the Bartlett's test of sphericity was significant (χ^2 = 16250.18; *df* = 903; $p < .05$). Both of these results indicate that factors can be extracted from the items. Then, the scree plot including the parallel analysis result was obtained as shown in Figure 1.

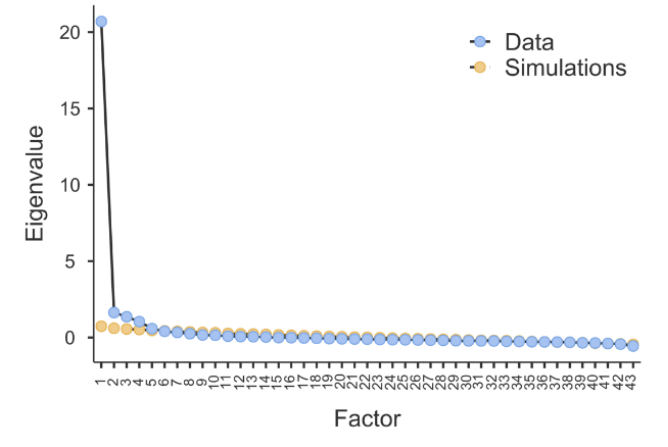


Figure 1. *Scree plot of the initial EFA.*

As can be seen in Figure 1, it is understood that the draft items can be grouped under four factors, whereas the items have a single dominant factor. Considering that the items would later be calibrated according to GRM, one of the unidimensional IRT models, it was decided to continue the analysis with a single-factor structure. When the unidimensional structure was tested, it was seen that 43 items could be gathered under a single factor explaining 48.1% of the total variance. This was considered as another finding supporting unidimensionality. Then, five items were removed from the analysis according to the factor loading and communality criteria. Accordingly, 38 items could be gathered under a single factor explaining 53.51% of the total variance. The final EFA findings are presented in Table 1.

Table 1. *EFA findings for teacher collaboration items.*

Item	Factor loading	Uniqueness	Item No	Factor loading	Uniqueness
23	.83	.31	29	.73	.46
18	.82	.32	27	.74	.46
13	.82	.32	3	.73	.46
21	.81	.34	32	.73	.47
11	.81	.34	5	.72	.48
16	.81	.35	9	.72	.48
22	.80	.36	1	.71	.49
6	.80	.37	20	.71	.50
14	.80	.38	2	.70	.51
19	.79	.40	24	.68	.54
10	.78	.43	34	.66	.57
41	.75	.44	40	.66	.57
12	.75	.44	33	.65	.58
30	.75	.44	38	.64	.58
7	.75	.44	42	.63	.60
17	.75	.44	37	.63	.61
4	.74	.45	15	.62	.62
39	.74	.45	25	.61	.62
43	.74	.45	8	.55	.70

Note. The first factor explained 53.51% of the variance. KMO = .97; $\chi^2(703) = 15532.72$, $p < .05$. Cronbach's α = .97; McDonald's ω = .98.

According to the EFA results, the minimum factor loading has been found as .55; and the maximum uniqueness (therefore minimum communalities) value has been found as .70. Internal consistency of the items is also very high (Cronbach's alpha = .97; McDonald's Omega = .98), which means that the items are functioning very consistently.

IRT calibration was performed with the remaining 38 items as a result of EFA. For this purpose, firstly, the assumption of local independence was tested, and then the item-model fit was examined with the S_χ^2 statistic. Accordingly, six items that violated local independence and 14 items that did not meet the item-model fit ($p < .05$), a total of 20 items, were removed from the data set. Even a high number of items were removed from the draft form of the scale, to avoid causing a content validity issue, remain items were carefully investigated according to the teacher collaboration construct. Due to intentionally developing three times as many items to the draft form, we decided that it did not cause any content validity problem. As a result, IRT calibrations were performed with the remaining 18 items. The results of item parameters and item-model fit statistics for the remaining 18 items are given in Table 2.

Table 2. *Item Parameters and item-model fit statistics for teacher collaboration items.*

Item No	*a* [SE]	*b1* [SE]	*b2* [SE]	*b3* [SE]	*b4* [SE]	*RMSEA* S_χ^2	*p* S_χ^2
	2.262	NA	-2.572	-1.183	.135	.018	.241
2	[.193]		[.215]	[.097]	[.075]		
3	2.530	NA	-2.829	-1.454	-.142	.000	.808
	[.222]		[.262]	[.107]	[.071]		
5	2.291	NA	-2.422	-1.191	.003	.000	.640
	[.195]		[.193]	[.095]	[.073]		
9	2.973	NA	-2.912	-1.728	-.616	.000	.890
	[.297]		[.300]	[.121]	[.070]		
11	4.367	NA	-2.646	-1.629	-.482	.026	.159
	[.456]		[.228]	[.103]	[.062]		
13	4.474	NA	-2.586	-1.583	-.486	.000	.758
	[.461]		[.227]	[.100]	[.062]		
16	3.584	-2.935	-2.359	-1.229	-.312	.018	.262
	[.328]	[.327]	[.177]	[.086]	[.064]		
18	3.457	NA	-2.080	-1.149	.021	.000	.816
	[.298]		[.145]	[.082]	[.066]		
20	2.697	-2.781	-2.190	-1.275	-.070	.000	.968
	[.228]	[.247]	[.158]	[.093]	[.069]		
22	3.773	NA	-2.398	-1.325	-.275	.014	.345
	[.350]		[.184]	[.089]	[.063]		
23	4.777	NA	-2.470	-1.407	-.385	.000	.708
	[.491]		[.192]	[.089]	[.061]		
24	1.960	-2.876	-1.544	-.618	.408	.025	.082
	[.161]	[.265]	[.122]	[.081]	[.083]		
29	2.698	-3.351	-2.206	-1.245	-.012	.000	.761
	[.226]	[.422]	[.161]	[.092]	[.070]		
30	3.067	NA	-2.615	-1.543	-.163	.024	.151
	[.276]		[.225]	[.107]	[.067]		
32	2.623	NA	-2.799	-1.693	-.173	.000	.656
	[.236]		[.255]	[.122]	[.069]		
37	1.815	-3.541	-2.899	-1.565	-.196	.019	.230
	[.169]	[.405]	[.270]	[.130]	[.080]		
41	2.487	-3.468	-2.306	-1.132	.141	.012	.369
	[.208]	[.455]	[.175]	[.090]	[.073]		
43	2.758	-3.280	-2.573	-1.600	-.379	.014	.340
	[.252]	[.418]	[.214]	[.113]	[.069]		

As can be seen, there is no *b1* parameter for 11 of the 18 items. The reason for this is that the sufficient number of participants did not select the strongly disagree response category in these items for a valid estimation. In other words, it was observed that these items worked as if they

had four options. In addition, when the *b4* parameters are examined, it is understood that they are mostly close to zero. Therefore, individuals whose level of teacher collaboration is above the average can easily select the highest response category, strongly agree. On the other hand, item discrimination, *a* parameters were found very high (minimum value for a parameter is 1.815). When *b* and *a* parameters interpreted together, we can say that the items can discriminate the teachers whom has trait level below the average ($\theta < 0$) better. This also can be seen in the Option Response Functions (ORFs) which presented in Figure 2.

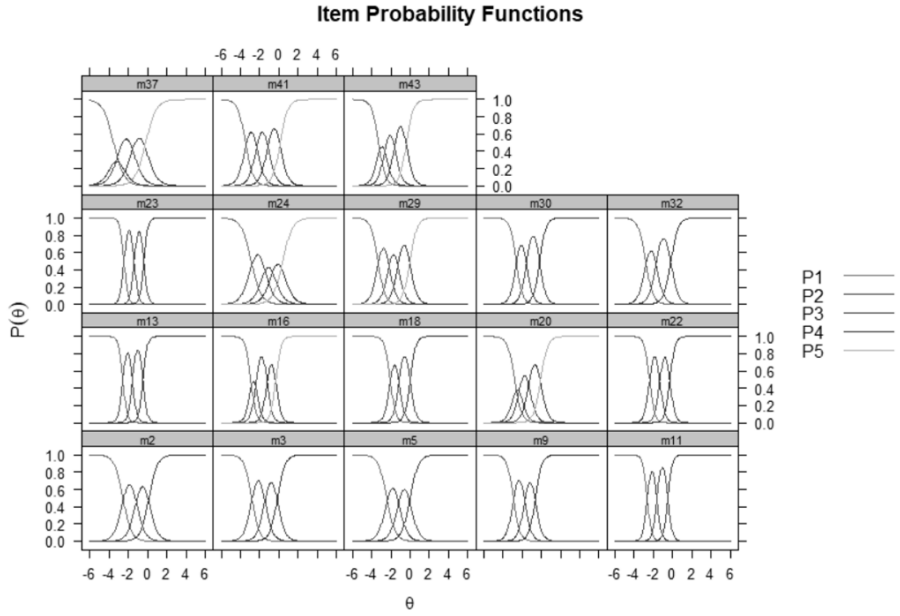


Figure 2. *ORFs for teacher collaboration items.*

The ORFs of 11 items with four response categories reflect this situation. In these 11 items, response functions belonging to only four options are observed. However, the ORFs are quite different from each other except for item 37. This means that the options for the items worked as expected. In item 37, the disagree option was dominated by the strongly disagree and neither agree nor disagree options. The item information functions for the remaining 18 items are shown in Figure 3. The test information function and standard error function obtained from the remaining 18 items are given in Figure 3.

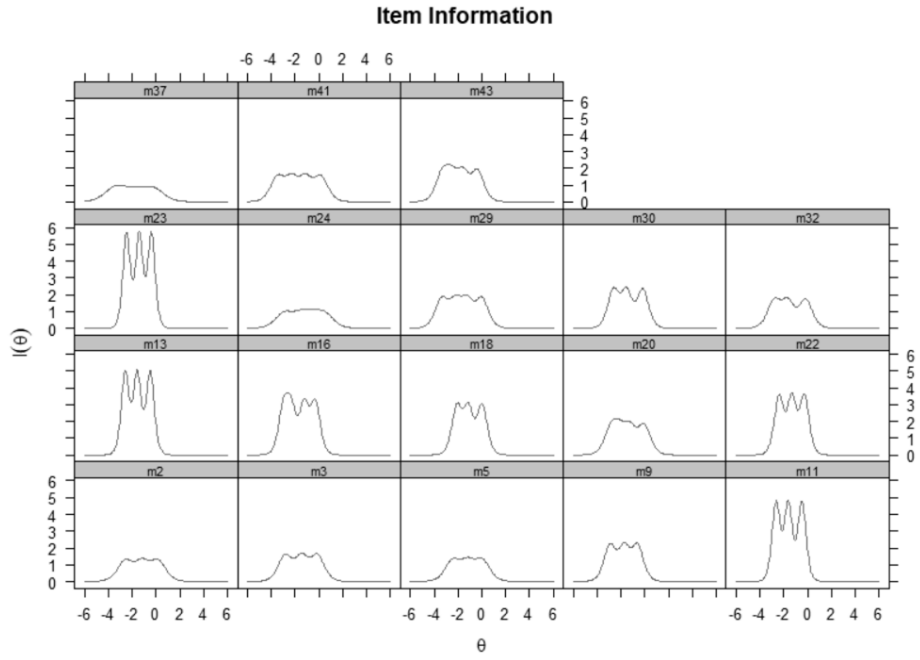


Figure 3. *IIFs for teacher collaboration items.*

When Figure 3 is analyzed, it is understood that the items mostly provided more information for individuals whose level of teacher collaboration was below average (θ < 0). Since the item information is directly related to item discrimination, these results in the information function were expected. A similar situation is reflected in the test information function in Figure 4.

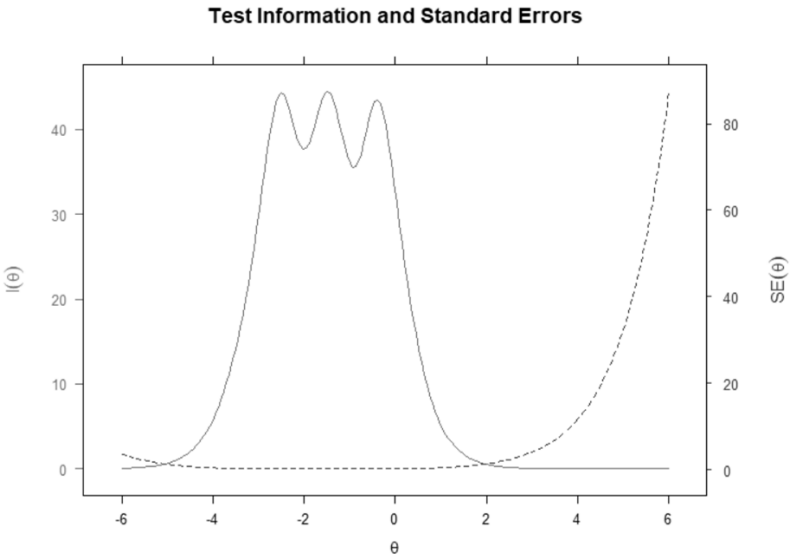


Figure 4. *TIF and SE Function of the teacher collaboration scale.*

When the test information function is analyzed, it is understood that the 18-item scale provides the most information for individuals between -3 and +1 trait level. Since the standard error function and the information function are inversely proportional, the scale also predicts with a lower standard error for individuals in this range. Similarly, the marginal reliability function was also obtained, and this function is presented in Figure 5.

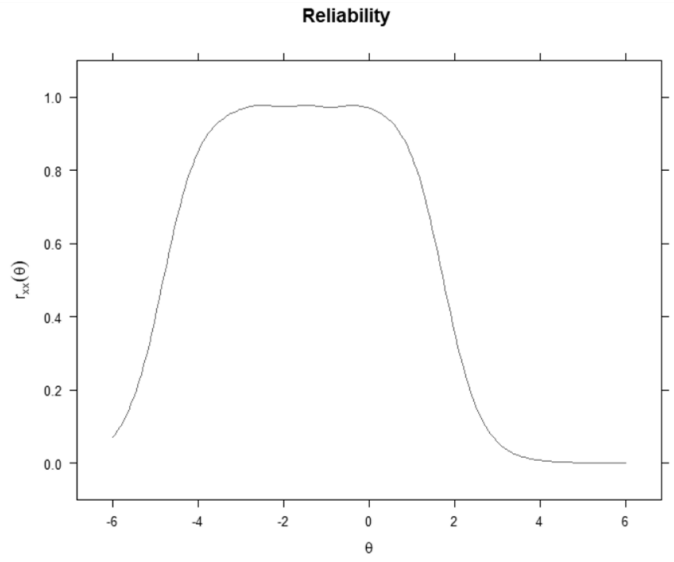


Figure 5. *Marginal reliability function of teacher collaboration scale.*

When Figure 5 is analyzed, similar to the test information function, it is understood that the scale is most reliable for individuals whose level of teacher collaboration is below the average. Finally, a Differential Item Functioning (DIF) study was conducted with Lord's chi-square method to determine whether the ORFs of the 18 items were invariant with respect to gender, and no DIF was detected in any item (p > .01). This shows that the scale items measure collaboration similarly for male and female teachers.

DISCUSSION and CONCLUSION

One of the important skills that 21st-century teachers should have is collaboration. In this sense, it was aimed to develop a qualified instrument to measure teachers' collaboration behaviors based on IRT, which provides the invariance properties of item parameters and individual characteristics. In accordance with the scale development procedures, a literature review was conducted, and existing scales and collaboration classifications were examined (Cerit, 2009; Child & Shaw, 2016; Çelebi *et al.*, 2016; Dumay *et al.*, 2013; Goddard *et al*, 2010; Gruenert *et al.*, 1998; Lench *et al.*, 2015; Özüdoğru, 2021; Saylık & Arastaman, 2022; Tschannen-Moran, 1998; Yılmaz & Çelik, 2020). When the scales directly related to teacher collaboration were examined, it was seen that Godard *et al.*'s (2010) scale had three dimensions: formal collaboration, frequency of collaboration in teaching, and teachers' collaboration on teaching policies; Dumay *et al.* (2013)'s teacher collaboration scale had four dimensions: "participation in school professional networks", "intellectual stimulation", "common vision", and "quality of professional relationships". It was noteworthy that none of these scales was structured on the basis of item-response theory, and this is a deficiency in the literature.

When the literature is examined, Gruenert (1998) tried to reveal the co-operation in school culture, Çelebi *et al.* (2016) tried to reveal the co-operation of teachers, Yılmaz and Çelik (2020) tried to reveal the attitude towards professional co-operation. Tschannen-Moran (1998) developed a scale to reveal teacher collaboration within the scope of his doctoral thesis in 1998. In the scale developed within the scope of this research, items were structured in line with current educational paradigms. At the same time, with this scale developed for teacher collaboration, an attempt was made to fill the gap in the domestic literature on teacher collaboration.

As a result of the study, a unidimensional scale consisting of 18 items to measure teacher collaboration was developed on the basis of GRM. When the parameters of the items were examined, it was seen that they ranged between 1.815 and 4.777. Therefore, all items in the scale are highly discriminative. On the other hand, when the ORFs of the items were examined, it was seen that although the items were developed in the form of a five-point Likert scale, in most cases they were answered in the form of a four-point Likert scale. Although the five-point Likert scale is recommended in the literature (Aybek & Toraman, 2022), Yaşar and Aybek (2019) stated that the seven-point Likert scale they developed functioned like a five-point Likert scale when ORFs were examined. A similar situation was observed in the current study.

The developed scale provides more information for individuals between -3 and +1 theta level and can measure quite reliably for this range. Therefore, it is seen that teachers with low level of cooperation can be better distinguished with this scale rather than teachers with high level of cooperation. Arias Gonzales (2015) and Yaşar and Aybek (2019), who developed an IRT-based scale to measure resilience, although not for collaboration, stated that their scales provided more information for individuals between -1.00 - 0.50 and -2.50 - 0.50, respectively. A similar situation was encountered in the current study.

In IRT-based scales, it is common for the test information to concentrate within a certain theta range, as item difficulty parameters typically cluster around lower and average levels of the latent trait. This pattern reflects the natural outcome of item development processes rather than a methodological flaw. In the current study, most items were more effective in distinguishing teachers with lower levels of collaboration, which may also be related to the tendency of collaboration behaviors to show limited variance at higher levels, leading to potential ceiling effects. To minimize this limitation, future studies may consider developing additional items that capture more sophisticated and high-level collaboration behaviors, incorporating items with higher discrimination parameters, and recalibrating the scale using samples that include more teachers with high collaboration levels.

Finally, a DIF study was conducted for the remaining 18 items in the scale and it was observed that none of the items showed DIF by gender. Accordingly, the scale items function similarly for both men and women.

The finding that the scale provides the highest level of reliability for individuals whose collaboration levels fall below the average is particularly noteworthy. This pattern indicates that the items are more effective in discriminating teachers who exhibit lower levels of collaboration, which may be a consequence of the distributional characteristics of the construct in real educational contexts. In many school settings, collaboration is an expected and socially desirable behavior; therefore, teachers often report higher levels of collaboration, creating reduced variability at the upper end of the latent continuum. Such ceiling tendencies naturally lead IRT-based scales to yield more information for individuals located in the lower regions of the trait. From a coverage perspective, this suggests that the scale is particularly sensitive in identifying teachers who require support or intervention regarding collaboration, which may be advantageous for practical applications such as needs assessment or program evaluation. However, it also indicates a relative limitation in distinguishing among individuals with high collaboration levels. To address this limitation, future research could develop more difficult items that capture advanced or sophisticated forms of collaboration (e.g., data-driven joint decision making, interdisciplinary instructional design, or long-term collaborative projects). Additionally, incorporating items with higher discrimination parameters and calibrating the scale with samples that include more highly collaborative teachers may help broaden the scale's effective measurement range.

This scale was found to better discriminate practitioners with low collaboration behavior; in future research, items can be added to the scale especially for practitioners with high levels of collaboration. Since some of the items in this study were found to function as a four-point Likert scale, the ORFs can be re-examined by applying this scale as a four-point Likert scale. In addition, the applicability of the items as CAT can be tested with post-hoc simulation and live applications. This study was conducted with teachers working in public schools in Denizli, Turkey. The fact that the sample consisted solely of public school teachers in Denizli limits the generalizability of the findings; therefore, this constitutes a limitation of the study. A similar study can be conducted in different provinces, countries and cultures and compared with the results of this study.

In addition to these findings, the present study offers an innovative contribution to the measurement of teacher collaboration. This is the first scale developed within the local context using an Item Response Theory framework, allowing for item-level precision, model-based reliability, and improved discrimination across different levels of collaboration. Therefore, the study provides a novel psychometric approach to understanding teacher collaboration and fills an important methodological gap in the literature.

Implications for Practice and Future Research

The findings of this study suggest several practical implications for educational settings. First, the scale's high sensitivity in distinguishing teachers with lower levels of collaboration indicates that it can serve as an effective diagnostic tool for school administrators and policymakers. By identifying teachers who require support in developing collaborative practices, schools can design targeted professional development programs, mentorship structures, or peer-support initiatives. Furthermore, since collaboration is a key component of school culture and instructional improvement, institutions may integrate the scale into regular school-wide evaluation processes to monitor collaboration levels over time. The unidimensional structure of the scale and its strong measurement precision across lower levels also make it suitable for use in needs assessment studies, intervention evaluations, and strategic planning efforts aimed at improving collaborative climates in schools.

The results also reveal important directions for future research. Since the scale provides more information for individuals with lower levels of collaboration, developing additional items that capture higher-level collaborative behaviors is recommended to extend the scale's effective measurement range. Future studies may also examine the scale using a four-point Likert structure, as ORF analyses showed that several items functioned similarly to a four-point format. Additionally, the applicability of the scale within computerized adaptive testing (CAT) environments can be explored through post-hoc simulations and live applications. Given that the current sample consisted solely of teachers working in public schools in Denizli, replicating the study with larger and more diverse samples across different regions, cultures, and educational systems would contribute to the generalizability and cross-cultural validity of the scale. Cross-national studies could also investigate whether collaboration exhibits similar latent structures in different contexts. Finally, future research may explore how teacher collaboration relates to instructional quality, student outcomes, professional identity, and school leadership practices.

Acknowledgments

We would like to express our gratitude to all the teachers who generously contributed to this research with their time and insights. Their participation was invaluable to the success of this study

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.
Ethics Committee Number: Ethical approval for this research was obtained from the Ethics Committee of Pamukkale University (Approval No: E-93803232-622.02-297694).

Data Availability Statement

Research data is available and can be sent with e-maill upon request.

Informed Consent

Written informed consent was obtained from the participants who participated in this study.

Contribution of Authors

Y.Ö.H: Investigation, Resources, Visualization, Software, and Writing-original draft. **E.C.A.:** Methodology, Formal Analysis, Supervision, and Validation. Authors may edit this part based on their case.

Orcid

Yeliz Özkan Hıdıroğlu https://orcid.org/0000-0002-5176-1235
Eren Can Aybek https://orcid.org/0000-0003-3040-2337

REFERENCES

Ahrari, S., Roslan, S., Zaremohzzabieh, Z., Mohd Rasdi, R., & Abu Samah, A. (2021). Relationship between teacher empowerment and job satisfaction: A Meta-Analytic path analysis. *Cogent Education, 8*(1), Article 1898737. https://doi.org/10.1080/2331186X.2021.1898737

American Management Association. (2019, January 24). AMA critical skills survey: Workers need higher level skills to succeed in the 21st century. https://www.amanet.org/articles/ama-critical-skills-survey-workers-need-higher-levelskills-to-succeed-in-the-21st-century/

Arias González, V.B., Crespo Sierra, M.T., Arias Martínez, B., Martínez-Molina, A., & Ponce, F.P. (2015). An in-depth psychometric analysis of the Connor-Davidson resilience scale: Calibration with Rasch Andrich model. *Health and Quality of Life Outcomes, 13*. https://doi.org/10.1186/s12955-015-0345-y

Aybek, E.C. (2021) *Data preparation for factor analysis.* URL https://shiny.eptlab.com/dp2fa/

Aybek, E.C. & Toraman, C. (2022). How many response categories are sufficient for Likert type scales? An empirical study based on the Item Response Theory. *International Journal of Assessment Tools in Education, 9*(2), 534-547. https://doi.org/10.21449/ijate.1132931

Barnard, C.I. (1958). Elementary conditions of business morals. *California Management Review*, *1*(1), 1-13. https://doi.org/10.2307/41165

Bertrand, L., Roberts, R.A., & Buchanan, R. (2006). Striving for success: Teacher perspectives of a vertical team initiative. *National Forum of Teacher Education Journal-Electronic*, *16*(3), 1-10.

Boeschen Hospers, J.M., Smits, N., Smits, C., Stam, M., Terwee, C.B., & Kramer, S.E. (2016). Reevaluation of the Amsterdam inventory for auditory disability and handicap using Item Response Theory. *Journal of Speech, Language, and Hearing Research, 59,* 373-383. https://doi.org/10.1044/2015_JSLHR-H-15-0156

Bursalıoğlu, Z. (2000). *Okul yönetiminde yeni yapı ve davranış [New construct and behavior in school administration]*. Pegem.

Çelebi, N., Vuranok, T.T. & Turgut, I.H. (2016). Reliability and validity of Teachers' collaboration level determination scale. *Kastamonu Education Journal, 24*(2), 803-820. https://izlik.org/JA94EP34MJ

Cerit, Y. (2009). The effects of servant leadership behaviors of school principals on teachers' job satisfaction. *Educational Management Administration & Leadership, 37*(5), 600-623. https://doi.org/10.1177/17411432093396

Chalmers, R.P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1-29. https://doi.org/10.18637/jss.v048.i06

Child, S.F., & Shaw, S. (2016). *Collaboration in the 21st century: Implications for assessment.* Cambridge Assessment.

Coke, P.K. (2005). Practicing What We Preach: An argument for cooperative learning opportunities for elementary and secondary educators. *Education, 126,* 392–398. https://eric.ed.gov/?id=EJ765674

de Ayala, R.J. (2009). *The theory and practice of item response theory*. The Guilford Press.

Dumay, X., Boonen, T., & Damme, J.V. (2013). Principal leadership long-term indirect effects on learning growth in mathematics. *The Elementary School Journal, 114*(2), 225-251. https://doi.org/10.1086/673198

Egodawatte, G., McDougall, D.E., & Stoilescu, D. (2011). The effects of teacher collaboration in Grade 9 Applied Mathematics. *Educational Research in Policy and Practice, 10*(3), 189–209. https://doi.org/10.1007/s10671-011-9104-y

Euwema, M., & Van der Waals, J. (2007). *Teams in scholen: Samen werkt het beter [Teamwork in schools: Together we're strong]*. Leusden: BMC.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3). 272-299. https://doi.org/10.1037/1082-989X.4.3.272

Goddard, Y.L., Goddard, R.D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record*, *109*(4), 877-896.

Goddard, Y.L., Miller, R., Larson, R., Goddard, R., Madsen, J., & Schroeder, P. (2010). *Connecting principal leadership, teacher collaboration, and student achievement.* Paper prepared for the AERA 2010 Annual Meeting. Texas A & M University and University of Michigan.

Gruenert, S.W. (1998). *Development of a school culture survey* [Unpublished digital dissertation]. University of Missouri-Columbia Missouri. http://www.MLLC.org

Khasawneh, Y.J.A., Alsarayreh, R., Ajlouni, A.A.A., Eyadat, H.M., Ayasrah, M.N., & Khasawneh, M.A.S. (2023). An examination of teacher collaboration in professional learning communities and collaborative teaching practices. *Journal of Education and E-Learning Research, 10*(3), 446–452. https://doi.org/10.20448/jeelr.v10i3.4841

Lench, S., Fukuda, E., & Anderson, R. (2015). *Essential skills and dispositions: Developmental frameworks for collaboration, communication, creativity, and self-direction.* Center for Innovation in Education at the University of Kentucky.

Meyer, A., Richter, D., & Hartung-Beck, V. (2020). The relationship between principal leadership and teacher collaboration: Investigating the mediating effect of teachers' collective efficacy. *Educational Management Administration & Leadership. Online first.* https://doi.org/10.1177/1741143220945698

Milli Eğitim Bakanlığı (MEB). (2017). Öğretmenlik mesleği genel yeterlikleri [General qualifications of the teaching profession]. http://oygm.meb.gov.tr/meb_iys_dosyalar/2017_12/11115355_YYRETMENLYK_MESLEYY_GENEL_YETERLYKLERY.pdf

National Research Council of the National Academies. (2012). *Research universities and the future of America* (pp. 1-28). Washington.

Organisation for Economic Co-operation and Development (OECD). (2018). *The future of education and skills: OECD Learning Framework* 2030. https://www.oecd.org/content/dam/oecd/en/publications/reports/2018/06/the-future-of-education-and-skills_5424dd26/54ac7020-en.pdf

Özkan Hıdıroğlu, Y., & Tanrıöğen, A. (2020). Development of Teachers' Empowerment Scale: A validity and reliability study. *International Journal of Assessment Tools in Education, 7*(4), 753-772. https://doi.org/10.21449/ijate.693398

Özkan Hıdıroğlu, Y. & Tanrıöğen, A. (2021). Examining teachers' perceptions on empowerment according to various variables. *International Journal of Curriculum and Instructional Studies, 11*(1), 127-146. https://doi.org/10.31704/IJOCIS.2021.007

Partnership for 21st Century Skills. (2009). A. framework for 21st century learning. Tucson: AZ: P21. www.21stcenturyskills.org

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Richter, E., Fütterer, T., Meyer, A., Eisenkraft, A., & Fischer, C. (2022) Teacher collaboration and professional learning. *Examining professional development during a national education reform - In: Zeitschrift für Pädagogik, 68* (6), 798-819 https://doi.org/10.25656/01:29309

Roschelle, J., & Teasley, S.D. (1995) The construction of shared knowledge in collaborative problem solving in C O'Malley (ed) *Computer supported collaborative learning*. Springer-Verlag, Berlin, 69–97.

Samejima, F. (1996). *Polychotomous responses and the test score*. Tennessee: The University of Tennessee.

Saylık, N., & Arastaman, G. (2022). Developing the school collaboration scale validity and reliability scale. *Van Yuzuncu Yil University Journal of Education, 19*(2), 462-483. https://doi.org/10.33711/yyuefd.1079626

Szczesiul, S., & Huizenga, J. (2014). The burden of leadership: Exploring the principal's role in teacher collaboration. *Improving Schools, 17,* 176–191. https://doi.org/10.1177/1365480214534545

The jamovi project (2022). *Jamovi*. (Version 2.3) [Computer Software]. https://www.jamovi.org

Tschannen-Moran, M. (1998). *Trust and collaboration in urban elementary schools* [Unpublished doctoral dissertation]. The Ohio State University.

Wigglesworth, M. (2011). *The effects of teacher collaboration on students' understanding of high school earth science concepts* [Unpublished master's thesis]. Montana University, USA.

Wren, D.A., & Bedeian, A.G. (2009). *The evolution of management thought.* John Wiley & Sons.

Wullschleger, A., Daly, A.J., van Halem, N., Merki, K.M., & Rechsteiner, B. (2025) Collaboration on school improvement under different educational accountability systems in two countries. *Educ Asse Eval Acc 37,* 303–329. https://doi.org/10.1007/s11092-025-09460-1

Yaşar, M. & Aybek, E.C. (2019). A resilience scale development for university students: Validity and reliability study based on Item Response Theory. *Elementary Education Online, 18*(4), 1687-1699. https://doi.org/10.17051/ilkonline.2019.635031

Yılmaz, K. & Çelik, M. (2020). Development of Scale of Attitude towards Professional Collaboration among Teachers. *MANAS Journal of Social Studies, 9*(2), 731-740. https://doi.org/10.33206/mjss.584856

APPENDICES

Appendix 1. *Öğretmen İşbirliği Ölçeği*.

Lütfen aşağıdaki ifadelere ne ölçüde katıldığınızı belirtiniz.

No	Item	(1) Hiç Katılmıyorun	(2)	(3)	(4)	(5) Tamamen Katılıyorum
1	(2) Öğretmenler çalışmalarımı takdir ederler.					
2	(3) Öğretmenlerle işbirliği yapma fırsatına sahibim.					
3	(5) Öğretmenlerle projeler üretmek için işbirliği yaparım.					
4	(9) Öğretmenlerle kaynak, belge, …vb. şeyler paylaşırım.					
5	(11) Öğretmenlerle deneyim ve fikirlerimi paylaşırım.					
6	(13) Öğretmenlerle öğretimi iyileştirmek için işbirliği yaparım.					
7	(16) Öğretmenlerle mesleki gelişim faaliyetleri konusunda haberleşirim.					
8	(18) Öğretmenlerle işbirliği yaparken eğlenceli ve üretken zaman geçiririm.					
9	(20) Öğretmenlerle onların yaşadıkları problemler üzerine tartışırım.					
10	(22) Farklı branşlardan öğretmenler ile işbirliği yaparım.					
11	(23) Öğretmenlerle önemli mesleki bilgileri paylaşırım.					
12	(24) Öğretmenlerle birlikte çalışma sayfaları hazırlarım.					
13	(29) Mesleki sorunlarım varsa iş arkadaşlarım beni dinler.					
14	(30) Öğretmenlerin deneyimlerini kendi öğretimimde kullanabilirim.					
15	(32) Öğrencilerin öğrenme süreçleri işbirliği yoluyla kolaylaştırılır.					
16	(37) Okulumuzda öğretmenler arasında arkadaşlıklar kurulmuştur.					
17	(41) Çalışmalarımızda orta bir anlayışta buluşabiliriz.					
18	(43) Öğretmenlerin işbirliği içerisinde olması kendimi güvende hissetmemi sağlar.					

Appendix 2. *Teacher Collaboration Scale.*

Please indicate the extent to which you agree with the following statements.

No	Item	(1) Strongly Disagree	(2)	(3)	(4)	(5) Strongly Agree
1	Teachers appreciate my work.					
2	I have the opportunity to collaborate with other teachers.					
3	I collaborate with teachers to produce projects.					
4	I share resources/documents/...etc. with other teachers.					
5	I share my experiences and ideas with other teachers.					
6	I collaborate with teachers to improve teaching.					
7	I keep teachers informed about professional development activities.					
8	I have fun and productive time while collaborating with teachers.					
9	I discuss the problems we encounter with other teachers.					
10	I collaborate with teachers from different subject areas.					
11	I share important professional information with teachers.					
12	I prepare worksheets together with teachers.					
13	If I have a professional problem, my colleagues listen to me.					
14	I use the experiences of my colleagues in my own teaching.					
15	Collaboration simplifies school processes.					
16	A trust-based relationship is built among teachers.					
17	We can reach common decisions in our work.					
18	Having collaboration with teachers makes me feel secure.					