Article

Collaboration and Shared Responsibility in Team Teaching: A Large-Scale Survey Study

Aron Decuyper [1],*, Hanne Tack [1,2], Bénédicte Vanblaere [3], Mathea Simons [2,4] and Ruben Vanderlinde [1]

1   Department of Educational Studies, Ghent University, 9000 Ghent, Belgium; hanne.tack@ugent.be (H.T.);
    ruben.vanderlinde@ugent.be (R.V.)
2   Department of Social Sciences, University of Antwerp, 2000 Antwerp, Belgium;
    mathea.simons@uantwerpen.be
3   Artevelde University of Applied Sciences, 9000 Ghent, Belgium; benedicte.vanblaere@arteveldehs.be
4   School of Education, University of Antwerp, 2000 Antwerp, Belgium
*   Correspondence: aron.decuyper@ugent.be

Abstract: The practice of team teaching—how teachers deliver team teaching in the classroom—substantially determines its effect. The collaboration between the teachers and the level of shared responsibility between them are two important dimensions of the team teaching practice. To date, no instrument exists to measure these dimensions. However, in view of empirical research within the context of team teaching, such an instrument is important. Therefore, the Collaboration and Shared Responsibility in Team Teaching (CSTT) scale is developed, making it possible to assess these two important dimensions. The CSTT scale was used in a large-scale cross-sectional survey study (n = 555). Next to a validation of the scale, this study provides empirical evidence on the differences between groups of teachers regarding (a) teaching experience, (b) education type, and (c) frequency of team teaching. Results show that teachers overall report high scores on both dimensions. Further, this study indicates that there are no significant differences between the groups based on (a) teaching experience and (b) education type for both collaboration and shared responsibility. There are, however, significant differences between groups in terms of the (c) frequency of team teaching.

Keywords: team teaching; measurement instrument; collaboration; shared responsibility; factor analysis; measurement invariance

1. Introduction

Until now, teaching has remained a highly individualised activity, with only little collaboration with other teachers [1]. Teachers seem to work primarily in their own classroom, largely isolated from other colleagues [2]. This isolation of teachers may impair learning opportunities for teachers as well as for students [3]. In this respect, evidence suggests that teacher collaboration results in positive outcomes for both teachers and students [4]. Teachers who collaborate can feel less isolated [5] and be more effective in their teaching [6,7]. Further, students whose teachers collaborate may experience higher learning outcomes, richer and more varied lessons, and increased support [8]. Hence, collaboration has been put forward by scholars as a way to improve teachers' teaching practice (e.g., [6,9,10]).

Collaboration within schools has, therefore, gained importance [11]. In this respect, educational institutions show a growing interest in teaching models in which teachers are more committed to collaborating, sharing expertise and experiences, supporting each other, and learning collaboratively [12], such as team teaching [13]. Team teaching can be described as two or more teachers in some level of collaboration in the planning, delivery, and/or evaluation of a course or courses [14]. Given its promising character [15], attention for team teaching has increased significantly during the past two decades [16]. This applies

to both fundamental research and educational practice [17]. Despite an increased interest in and emphasis on teacher collaboration [18], team teaching has only been studied to a limited extent [19].

The practice of team teaching—how teachers deliver team teaching in the classroom—substantially determines its effect [20]. Because the practice of team teaching plays a crucial role in the impact of team teaching, Sweigart and Landrum [21] recommend that further analysis of the practice of team teaching is necessary. However, to date, there exists no appropriate measurement instrument that captures important dimensions of the practice of team teaching. Therefore, this study attempts to contribute to the research base on team teaching by developing an instrument to assess important dimensions of the practice of team teaching (i.e., collaboration and shared responsibility). By pioneering the development of such an instrument, this study aims to fill an existing gap. Additionally, this study aims to investigate whether differences exist between groups of teachers regarding these important dimensions. It thus goes beyond instrument development and also advances research on collaborative learning environments. In this way, it not only fills a gap in research but also reveals valuable insights that can foster more effective team teaching practices.

2. Theoretical Framework

2.1. Team Teaching

In the literature, there are many synonyms (e.g., co-teaching, collaborative teaching, and cooperative teaching) for team teaching that are often used interchangeably [19,22]. In this regard, the concept of team teaching is considered to be an umbrella term [19]. It generally refers to the collaboration between two or more teachers in the planning, delivery, and/or evaluation of a course or courses [14]. Additionally, various definitions of team teaching can be found [16,19,23]. For instance, Welch, Brownell [24] (p. 38) define team teaching as: "the simultaneous presence of two educators in a classroom setting who share responsibility in the development, implementation, and evaluation of direct service in the form of an instructional or behavioural intervention to a group of students with diverse needs". Thousand, Villa [25] (p. 5) describe it as: "when two or more people share responsibility for teaching some or all of the students assigned to a classroom". Fuller and Bail [26] define team teaching as two or more people sharing responsibility for teaching some or all of the students assigned to a class. Overall, these definitions highlight the collaborative nature of team teaching, as well as the shared responsibility among the teachers involved.

Although there is some research, theory development on team teaching is still in its infancy [19], especially within the context of compulsory education. Furthermore, research is mainly small-scale and qualitative in nature [21,27], and has an almost sole focus on experiences and perceptions about team teaching. Despite these gaps, team teaching is regarded in the literature as a teaching model that holds several benefits for both teachers and students. Teachers report increased emotional and professional support, increased reflective dialogue, professional and personal growth, and learning gains [23]. Nevertheless, concerns have been recognised as well. For instance, teachers using team teaching indicate that compared to individual teaching the workload increases [28,29]. Moreover, they perceive that students compare between teachers and that this could potentially even lead to competition [30]. The literature also mentions benefits for students. Students who are taught in a team teaching environment report richer and more varied learning opportunities, quicker assistance, and more individualised attention [23]. However, some concerns also exist for students. For instance, students may potentially be confused when being confronted with multiple teachers in the classroom and not know which teacher they should turn to [31].

2.2. Measurement Instruments

Although until now, no appropriate instrument exists that advances insights in teachers' team teaching practice, several measurement instruments have been developed to

capture aspects related to one's team teaching practice. These instruments have either a focus on learners' perceptions of team teaching [32], on student teachers' perceptions on team teaching [33], or on perceived advantages and disadvantages of team teaching [34]. The existing measurement instruments mainly respond to experiences and perceptions, but do not so much try to capture typical features of the practice of team teaching itself, which is the aim of this study. The newly developed instrument could be used to complement the existing instruments.

2.3. The Practice of Team Teaching

The practice of team teaching—how teachers deliver team teaching in the classroom—is expressed in the literature through the models of team teaching [25,35]. The models of team teaching represent the ways in which team teaching is established in the classroom (e.g., observation model, parallel model, teaming model). However, these team teaching models do not correspond to the complex reality within the classroom. For this reason, the objective of this study is to develop an instrument able to capture collaboration and shared responsibility as two important dimensions of the practice of team teaching. Simons, Coetzee [22] recently placed the most common models of team teaching on two continua: collaboration and shared responsibility.

Following Vangrieken, Dochy [4], collaboration can be defined as "joint interaction in the group in all activities that are needed to perform a shared task". It refers to teachers actually doing things together [36]. Doing things together entails negotiation, discussion, and consideration of opposing viewpoints [37]. Following research by Valckx, Devos [38], shared responsibility means that colleagues create a common sense of responsibility for all students' learning [39]. Shared responsibility is, according to Sleegers, Den Brok [40], part of interpersonal capacity. This notion is echoed by Griffin and Robertson [41], who shift responsibility for a class from the teacher to the team in which that teacher works. More specifically, it represents a shift from the traditional approach which sees a teacher mainly responsible for his or her own class, to one where the whole team takes responsibility for the learning of the students in the classes for which they cover [42]. To accomplish this, team members must take on additional work, support other teachers, and contribute to the decisions that will support students in other classes [42].

3. Research Goals

The main aim of this study is to provide empirical evidence on important dimensions of the practice of team teaching (i.e., collaboration and shared responsibility). Presumably, high levels of collaboration and shared responsibility will be assessed, due to the inherent nature of team teaching. However, there is, to the best of our knowledge, no appropriate instrument available to capture these important dimensions. Therefore, it is necessary to first develop an instrument to measure these important dimensions. It is hypothesised that it is possible to develop such an instrument. Next, this study aims to investigate differences between groups of teachers regarding these dimensions of the practice of team teaching. More specifically, differences between groups of teachers based on (a) teaching experience, (b) education type, and (c) frequency of team teaching. Following [22], it is assumed that the degree of collaboration and shared responsibility depends on the practice of team teaching. However, there is no empirical research that demonstrates in what ways this occurs. This study aims to fill this research gap and thus contribute to increasing knowledge in this area. In addition, this study provides valuable insights that can inform practice and policy regarding team teaching.

In accordance with the purposes of this study, three research goals (RG) are formulated:

- RG1: Development of an instrument to capture collaboration and shared responsibility in team teaching;
- RG2: Providing empirical evidence on collaboration and shared responsibility in the practice of team teaching;

- RG3: Investigating whether differences exist between groups of teachers regarding important dimensions of the practice of team teaching.

4. Method

The threefold purpose of this study is similarly reflected in the method. First, an instrument to measure collaboration and shared responsibility in team teaching was developed using four phases. Second, empirical evidence is provided on these two dimensions of the practice of team teaching. Third, differences in teachers' practice of team teaching across several groups of teachers (i.e., based on (a) teaching experience, (b) education type, (c) and frequency of team teaching) were explored, using two-sample *t*-tests and one-way analysis of variance (ANOVA), after tests for measurement invariance.

4.1. Development of an Instrument to Capture Collaboration and Shared Responsibility in Team Teaching (RG1)

To address the first research goal, a measurement instrument was developed and descriptive statistics were used. The instrument (i.e., the CSTT scale) was developed in four phases, a procedure inspired by Gehlbach and Brinkworth [43]. First, based on the research literature on team teaching, a list of items for each dimension (i.e., collaboration and shared responsibility) was generated. Second, an expert review was conducted with a panel of 12 teachers, teacher educators, and researchers. Third, a pilot study was conducted with 20 team teachers in order to improve content validity. Fourth, a validation and reliability study was carried out with 555 teachers, based on exploratory factor analyses, confirmatory factor analyses, and internal consistency analyses.

4.1.1. Phase 1—Preliminary Version

The preliminary version of the CSTT scale consisted of 20 items organised in two scales: collaboration (12 items) and shared responsibility (8 items). For the dimension of collaboration, five items were based on items of the Student Teachers' Team teaching Perceptions Questionnaire (STTPQ) from De Backer, Simons [33]. Seven items were inspired by conditions for successful collaboration, derived from the literature review about student teachers' team teaching by [14]. For the dimension of shared responsibility, four items were based on items of an instrument developed by Vangrieken, Grosemans [44]. Four additional items were newly constructed to capture other key tasks in the classroom context, such as evaluating the lesson, but also school outcomes of students (i.e., learning outcomes, well-being, and motivation). A 5-point Likert scale was used, ranging from 'I totally disagree' (0) to 'I totally agree' (4).

4.1.2. Phase 2—Expert Review

An expert review of the items was conducted to assess the content validity of the survey by requesting detailed responses concerning clarity, relevance, and quality of items. The expert panel consisted of 12 experts from the field of education (1 teacher, 4 teacher educators, and 7 researchers). One item (i.e., 'During team teaching classes, I feel that my team teaching colleague(s) and I work together efficiently.') was deleted within the scale of collaboration, because the item was too general and was covered by several other items. Furthermore, one item of the scale of shared responsibility was reworded to enhance item clarity (i.e., 'reflection' instead of 'evaluation'). Lastly, the sequence of items was modified; the last five items of the scale for collaboration were placed first, because they are more straightforward items to answer.

4.1.3. Phase 3—Pilot Study

After the measurement instrument was developed and reviewed by experts, a pilot study was conducted. The pilot study was undertaken over a three-week period (from 16 March to 6 April 2022) with 20 team teachers. Out of the 20 participants, 17 were female. The mean age of the participating team teachers was 35 years old (standard deviation

(SD) = 6.6; Range: 23–44) and they had on average 9 (SD = 6.3; Range: 0–20) years of experience in teaching. Two teachers worked in pre-primary education, eight in primary education, eight in secondary education, and two in adult education. Participants were asked to fill out the instrument, and afterwards they could leave written remarks. No major changes were made after the pilot study. The instrument so far consisted of 11 items for collaboration and 8 items for shared responsibility.

4.1.4. Phase 4—Validation and Reliability Study

A total of 555 participants in 86 Flemish (the Dutch speaking part of Belgium) schools completed the survey. The survey was conducted from March 2022 to June 2022. Data were collected by a convenience sample procedure. All Flemish schools, encompassing preschool, primary, secondary, and adult education, were contacted by e-mail with information about the purpose and the design of the study and asked to participate. In schools that agreed to participate in the study, the survey was administered to all teachers with team teaching experience. Teachers who indicated that they never engaged in team teaching during a course were not included in this study. The online platform Qualtrics was used and informed consent was obtained from all participants. The data was subjected to a rigorous cleaning process within the R statistical environment. Cases in which participants were missing data for the collaboration dimension (11 items) and the shared responsibility dimension (8 items) were identified and subsequently excluded from the analysis. This resulted in the removal of 101 participants from the analyses.

Most participants are female (85%), have a bachelor degree (86%), and are working full-time (70%) as a teacher. Their age differs from 22 to 62 years, with a mean age of 39 years (SD = 10.37). The 555 participants have a mean experience of 15 years (SD = 10.67), ranging from 0 to 41 years. More details are shown in Table 1.

Table 1. Descriptive statistics of the participants (n = 555).

Participants'	N (%)
Gender	
Male	84 (15.14)
Female	470 (84.68)
X	3 (00.18)
Degree	
Secondary education	3 (00.54)
Bachelor	475 (85.59)
Master	76 (13.69)
Doctorate	1 (00.18)
Employment	
Full-time	389 (70.09)
Part-time: 50% or more	164 (29.55)
Part-time: less than 50%	2 (00.36)
Education type	
Pre-primary	100
Primary	231
Secondary	156
Adult	53
Participants'	**Mean (SD)**
Age	38.81 (10.24)
Experience	
As teacher	14.38 (10.61)
In school	11.17 (10.28)

Data of 555 participants were used to validate the measurement instrument. Data were analysed by conducting (1) exploratory factor analyses (EFA) to examine the factor structure (i.e., number of factors), (2) confirmatory factor analyses (CFA) to assess the stability of the factor structure, and (3) reliability analyses—based on Cronbach's alpha—to determine the internal consistency of the factors. These subsequent steps of scale construction are based on recommendations of the American Educational Research Association, the American

Psychological Association, and the National Council on Measurement in Education (AERA, APA, NCME) [45].

To check the normality, the skewness and kurtosis of each item was screened [46,47] (see Appendix A; C = Collaboration; SR = Shared Responsibility). According to Kline [46], there may be a concern for non-normality if the absolute values of skewness are greater than three and if the absolute values of kurtosis are greater than ten. This is not the case, thus, the data in the present study could be considered as normal. However, based on visual inspection of the distribution, some items tended to differ from the normal distribution. Therefore, an estimator was used that calculates robust standard errors. In this study, the weighted least squares were used. The correlations between the items are positive and range from 0.17 (i.e., C11–CS3) to 0.86 (i.e., CS7–CS8), measuring that one item tends to increase when the other increases.

The Kaiser–Meyer–Olkin coefficient of sampling adequacy [48] and the Bartlett's test of Sphericity [49] were used to assess suitability of the data for factor analysis. The Kaiser–Meyer–Olkin coefficient for this dataset is 0.94, exceeding the recommended value of 0.60 [50]. Additionally, all the KMO values for the individual items are >0.86. The Bartlett test reached statistical significance (X^2 = 8186.78, df = 171, p < 0.001). Both measures indicate that properties of the correlation matrix justified factor analysis being carried out.

In order to minimise the chance of overanalysing the data, the total sample (n = 555) was randomly divided into two equal groups using the odds and evens split method [47]. More specifically, the dataset was split up in two random subsamples in which all team teachers were equally represented in order to create a development (n1 = 277) and a validation sample (n2 = 278). Thus, EFA and CFA, respectively, could be carried out on separate subsamples. Two-sample t-tests were conducted to examine possible significant differences between the two subsamples. Both subsamples were equivalent with regard to gender (i.e., male and female) (t(552) = −0.710, p = 0.478), age (t(552) = 1.083, p = 0.279), and teaching experience (t(553) = 0.834, p = 0.405).

The EFA was conducted on the data of the first subsample (n1 = 278) to identify the number of latent variables underlying the measured items without strong theoretical assumptions on how many factors existed [50]. Because team teachers' collaboration and shared responsibility tend to correlate [22] and the preliminary analysis confirmed factor dependence, an exploratory weighted least squares factoring analysis was performed, with a direct oblimin rotation (oblique rotation technique) [50]. The latter allows factors to be correlated and produces estimates of correlations among factors [51]. In order to perform the EFA, the psych package [52] in R was used.

As advised by O'Connor [53], several statistical criteria were used to determine the number of factors to withhold. The Kaiser criterion [54] to retain eigenvalues bigger than one, and Cattell's scree test [55] were carried out. Since both criteria sometimes overestimate the number of factors to withhold [56], Horn's parallel analysis [57] and Velicer's Minimum Average Partial (MAP) technique [58,59] were also conducted.

In addition, item factor loadings were screened. Following the recommendations of Hair, Black [60], all items with loadings of less than 0.70 were excluded from further analyses. Furthermore, all items in the retained factor solution with strong cross-loadings on other factors (i.e., when the gap between the primary target loading and the cross-loading is smaller than 0.25) were also removed [47]. These items are affected by more than one factor and are thus deemed too intricate [47].

A confirmatory weighted least square mean and variance adjusted factoring analysis was performed on the data of the second subsample (n2 = 277) to assess the stability of the proposed factor structure of the earlier conducted EFA [50]. For the CFA, several fit indices were calculated to determine whether the proposed factor structure of the EFA fits the empirical data. In order to conduct the CFA, the lavaan package in R [61] was used.

The following fit indices were evaluated: the χ^2-test and the associated p-value, the χ^2/df ratio, the comparative fit index (CFI) [62], the Tucker-Lewis index (TLI) [63], the root mean square error of approximation (RMSEA) [64], and the standardised root mean square

residual (SRMR) [65]. Chi-Square value along with its *p*-value and degrees of freedom is reported but disregarded, due to limitations of the Chi-square test [66]. For the χ^2/df ratio, a value ≤ 3 determines an acceptable fit [67]. Furthermore, CFI and TLI scores ≥ 0.90 indicate adequate fit, while scores of ≥ 0.95 indicate a good fit [68]. Following Hu and Bentler [68], cut-off values of ≤ 0.06 and ≤ 0.08 for RMSEA and SRMR, respectively, indicate a good fit.

Cronbach's alpha was calculated on the data of the complete sample (n = 555) as a measure of internal consistency in order to determine the psychometric quality of the scales. Factors with a Cronbach's alpha of 0.80 are considered reliable [47]. In order to perform the reliability analyses, the psych package in R [52] was used.

4.2. Empirical Evidence on Collaboration and Shared Responsibility in the Practice of Team Teaching (RG2)

To gain a first empirical insight into the dimensions of collaboration and shared responsibility in the practice of team teaching, descriptive statistics were used. The mean scores of the dimensions, their standard deviation, and their correlation (Chronbach's alpha) were calculated. Furthermore, a paired samples *t*-test was conducted to assess if there was a significant difference between the mean scores of the dimensions. The psych package in R [52] was used.

4.3. Differences in Teachers' Practice of Team Teaching across Several Groups of Teachers (RG3)

To address the third research question, two-sample *t*-tests and one-way analysis of variance (ANOVA) were used, after tests for measurement invariance. Multiple group measurement invariance based on multigroup confirmatory factor analyses was performed to test whether the factor structure of the developed measurement instrument is invariant across (a) teaching experience, (b) education type, and (c) frequency of team teaching.

To verify whether the developed instrument measures the same constructs, with the same structure across (a) teaching experience, participants are divided into two groups based on their teaching experience (i.e., teachers with less than five years of experience, teachers with more than five years of experience). Five years is chosen as a turning point, as this is also common in previous research (e.g., [69,70]). Regarding (b) education type, participants of four types are surveyed: pre-primary, primary, secondary, and adult education. The third group characteristic studied was (c) frequency of team teaching, indicating the frequency the participant uses team teaching. For this characteristic, the tipping point of once a week was opted for. Teachers who apply team teaching less than once a week and teachers who use team teaching more than once week are compared. At each step of the analysis, a series of factor models were estimated, which impose increasing constraints on the parameters for the groups. Before testing measurement invariance, a baseline model was determined for each group [71]. The two-factor structure of the CSTT scale served as the initial model that was tested when creating the baseline models for each group.

Four levels of measurement invariance are described (from less constrained to more constrained): (1) configural invariance, (2) metric invariance, (3) scalar invariance, and (4) strict invariance [72]. First, the baseline model was tested for equivalent factor structures (i.e., (1) configural invariance). It was tested by specifying the same measurement model across the groups. In this model, both the number of factors and the factor-indicator correspondence are the same, but all factor loadings and item intercepts are freely estimated within each group [46]. If only configural invariance is established, it would indicate that teachers conceptualise constructs (i.e., collaboration and shared responsibility) similarly, but it would not guarantee that individual items are interpreted in the same way. In the second model, the factor loadings were required to be equivalent across the groups [46], while the item intercepts were allowed to vary freely. Metric invariance (2) indicates that factor loadings are equal across groups and that items are, therefore, interpreted in a similar way. The third model constrains, additionally, the item intercepts to be equivalent across the groups [46]. If (3) scalar (or measurement) invariance is achieved, differences in means

of the observed items can be interpreted as a consequence of the differences in the means of the latent constructs. The (4) strict invariance model is the constrained version of the scalar model, where the factor loadings, intercepts, and residual variances are fixed across groups [73]. The multiple group measurement invariance is analysed using the multi-group CFA with weighted least square mean and variance adjusted [74]. The analyses are carried out in the lavaan package in R [61].

To compare the different models, changes in CFI, RMSEA, and SRMR were evaluated. According to Cheung and Rensvold [75] and Chen [76], ΔCFI should be smaller than or equal to 0.010 and ΔRMSEA smaller than or equal to 0.015. Additionally, Chen [76] also suggests a criterion of changes in ΔSRMR of 0.030.

When measurement invariance is established across (a) teaching experience, (b) education type, or (c) frequency of team teaching, it is possible to compare across these groups mean sum scores for collaboration and shared responsibility. To assess whether there are significant differences between these groups, two-sample *t*-tests and an ANOVA with Bonferroni post hoc tests were performed. When a significant difference was identified, effect sizes (Cohen's d effect size index (d) and Omega Squared (ω2)) were additionally calculated in view of interpreting the importance of the analysis results. All data analyses were conducted in R, using the psych [52] and rstatix package [77].

5. Results

5.1. Development of an Instrument to Capture Collaboration and Shared Responsibility in Team Teaching (RG1)

Varied statistical criteria were used to determine how many factors to retain: Kaiser criterion (two factors; eigenvalues of 10.170 and 2.201), Cattell's scree test (two factors) (see Figure 1), Horn's parallel analysis (two factors), and Velicer's MAP technique (original (1976) MAP Test is = two factors, revised (2000) MAP Test is = two factors). All criteria proposed a two-factor structure.
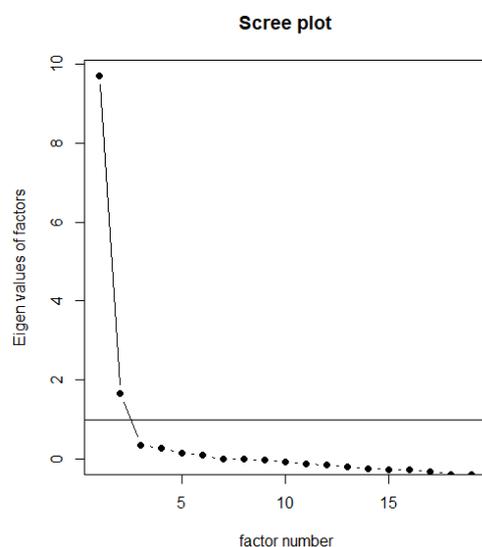
Figure 1. Cattell's scree test.

To interpret the factor structure, an exploratory weighted least squares factoring analysis (n1 = 278) with a direct oblimin rotation was applied. The item factor loadings of the two-factor solution, with the sum of squared loadings (SS) of 6.753 (Factor 1), and 4.167 (Factor 2) were examined. Of the 19 items, four have a factor loading of less than 0.70 (i.e., C10, SR2, SR7, and SR8). No item has a cross-loading of more than 0.25 with the other factor. Thus, 15 items were retained. The items that were deleted are shown in italics in Table 2.

A confirmatory weighted least square mean and variance adjusted factoring analysis was performed with the second subsample (n2 = 277) in order to confirm the number

of factors found in the EFA and determine whether they are independent or related to each other. A confirmatory weighted least square mean and variance adjusted factoring analysis was performed on the two-factor structure with 15 items resulting from the EFA (see Table 3).

Table 2. Loadings of items (n = 19) on the two rotated factors.

	Item	Factor 1	Factor 2
	During team teaching classes, I feel that my team teaching colleague(s) and I . . .		
C1	can rely on each other for questions and concerns.	0.729	
C2	complemented each other very well.	0.724	
C3	gave each other emotional support.	0.713	
C4	gave each other professional support.	0.770	
C5	mutually trust each other.	0.887	
C6	respect each other.	0.900	
C7	support each other.	0.835	
C8	openly discuss experiences.	0.798	
C9	are open to reflection.	0.810	
C10	*provide each other feedback.*	0.585	
C11	avoid competition.	0.713	
	During team teaching classes, my team teaching colleague(s) and I are responsible for . . .		
SR1	designing and preparing the lesson.		0.750
SR2	*classroom management.*		0.532
SR3	student evaluation.		0.718
SR4	teaching and supporting the class assignments.		0.765
SR5	evaluating the lesson.		0.818
SR6	students' learning outcomes.		0.824
SR7	*students' well-being.*		0.573
SR8	*students' motivation.*		0.619

Table 3. Loadings of items (n = 14) on the two rotated factors.

	Item	Factor 1	Factor 2
	During team teaching classes, I feel that my team teaching colleague(s) and I . . .		
C1	can rely on each other for questions and concerns.	0.830	
C2	complemented each other very well.	0.828	
C3	gave each other emotional support.	0.795	
C4	gave each other professional support.	0.857	
C5	mutually trust each other.	0.845	
C6	respect each other.	0.821	
C7	support each other.	0.895	
C8	openly discuss experiences.	0.796	
C9	are open to reflection.	0.845	
C11	avoid competition.	0.608	
	During team teaching classes, my team teaching colleague(s) and I are responsible for . . .		
SR1	designing and preparing the lesson.		0.760
SR3	student evaluation.		0.698
SR4	teaching and supporting the class assignments.		0.758
SR5	evaluating the lesson.		0.851
SR6	students' learning outcomes.		0.761
	Cronbach's Alpha	0.949	0.879

Several fit indices were calculated to determine whether the proposed factor structure of the EFA fits the empirical data: X^2 = 108.094, df = 89, p = 0.082; X^2/df = 1.213; CFI = 0.975; TLI = 0.971; RMSEA = 0.028; SRMR = 0.039, with all fit indices meeting the generally accepted norms for CFA [68]. The results of the CFA show a good fit for the initial two-factor model with collaboration (10 items) and shared responsibility (five items) as factors. Collaboration is defined as the joint interaction in the group in all activities that are needed to perform a shared task [4]. Shared responsibility means that colleagues create a common sense of responsibility for all students' learning [39].

A reliability analysis was performed on the complete dataset (n = 555) to examine the internal consistency of the two factors (i.e., collaboration and shared responsibility). The newly constructed scale is found to be highly reliable, with Chronbach's alphas of 0.949 and 0.879, respectively, for collaboration and shared responsibility as two important dimensions of the practice of team teaching (see Table 3).

5.2. Empirical Evidence on Collaboration and Shared Responsibility in the Practice of Team Teaching (RG2)

Descriptive statistics showed teachers reported high scores on both dimensions. The dimension collaboration (10 items) has a mean score of 3.54 (SD = 0.587) on a scale of 0 to 4, and for the dimension shared responsibility (five items), a mean score of 3.05 (SD = 0.880) was found. Since the correlation between both dimensions is 0.433, a high score for collaboration corresponds to a high score for shared responsibility, and vice versa. These results indicated that teachers experience a high degree of collaboration, and a high degree of shared responsibility as two related dimensions of the practice of team teaching. Additionally, a paired sample *t*-test demonstrated there is a significant difference between the mean score for collaboration and the mean score for shared responsibility (t(554) = 14.167, $p < 0.001$). Thus, teachers reported a significantly higher score for the dimension collaboration than for the dimension shared responsibility.

5.3. Differences in Teachers' Practice of Team Teaching across Several Groups of Teachers (RG3)

The current study attempted to establish scalar invariance for (a) teaching experience (i.e., teachers with less than five years of experience, teachers with more than five years of experience), (b) education type (i.e., pre-primary, primary, secondary, and adult education), and (c) frequency of team teaching (i.e., teachers who team teach less than once a week, teachers who team teach more than once a week).

Small changes in ΔCFI, ΔRMSEA, and ΔSRMR, and satisfying overall model results (see Appendix B) revealed measurement invariance for the models across (a) teaching experience [75,76]. All changes are smaller than 0.010 in ΔCFI, 0.015 in ΔRMSEA, and 0.030 in ΔSRMR [75,76], which means that teachers across the studied groups interpret the developed measurement instrument in a consistent manner. Furthermore, the criteria for measurement invariance across (b) education type are met, except for the ΔCFI (ΔCFI = 0.013 > 0.010) between the scalar and metric model [75,76]. However, the other criteria are met (ΔRMSEA = 0.004; ΔSRMR = 0.003) and, moreover, strict invariance (ΔCFI = 0.007; ΔRMSEA = 0.002; ΔSRMR = 0.001) is established [75,76]. Hence, it was opted to attribute measurement invariance across education type. Lastly, small changes in ΔCFI, ΔRMSEA, and ΔSRMR demonstrated measurement invariance between the models of the studied groups based on the (c) frequency of team teaching [75,76].

Based on the establishment of measurement invariance across (a) teaching experience, (b) education type, and (c) frequency of team teaching, it is possible to compare the mean sum scores across these groups.

The mean scores for collaboration and shared responsibility between teachers with less than five years of experience (n = 127, M_c = 3.51, M_{SR} = 2.98) and teachers with more than five years of experience (n = 428, M_c = 3.55, M_{SR} = 3.07) were compared, using a two-sample *t*-test. Results indicate no significant differences between the two groups for both collaboration (t(210.068) = -0.750, $p = 0.454$) and shared responsibility (t(212.881) = -0.963, $p = 0.337$). This means that teachers with less than five years of experience report the same extent of collaboration and shared responsibility in comparison with teachers with more than five years of experience.

Furthermore, one-way analyses of variance with Bonferroni post hoc tests were used to compare the mean scores between teachers of pre-primary (n = 100, M_c = 3.54, M_{SR} = 3.12), primary (n = 231, M_c = 3.49, M_{SR} = 2.98), secondary (n = 156, M_c = 3.62, M_{SR} = 3.06), adult education (n = 53, M_c = 3.55, M_{SR} = 3.18). Results indicate no significant differences among the different groups for both collaboration (F(3, 536) = 1.417, $p = 0.237$) and shared

responsibility (F(3, 536) = 1.021, *p* = 0.383). This means that teachers from pre-primary, primary, secondary, and adult education report the same extent of collaboration and shared responsibility.

Subsequently, the mean scores for collaboration and shared responsibility between teachers who team teach less than once a week (n = 148, M_c = 3.60, M_{SR} = 3.14) and teachers who team teach more than once a week (n = 407, M_c = 3.39, M_{SR} = 2.81) were compared, using a two-sample *t*-test. Results indicate significant differences between the two groups for both collaboration (t(216.279) = 3.252, *p* = 0.001, d = 0.35) and shared responsibility (t(237.401) = 3.737, *p* < 0.001, d = 0.38). This means that teachers who team teach less than once a week report a significantly lower score for collaboration and shared responsibility in comparison with teachers who team teach more than once a week. Moreover, Cohen's d effect size index indicates small differences between the two groups of teachers [78].

6. Discussion

The impact of team teaching is determined primarily by how it is put into practice [20]. In order to conduct further and more in-depth research on team teaching [21], it is necessary to have an instrument that can map that effective realisation of team teaching. Therefore, the Collaboration and Shared Responsibility in Team Teaching (CSTT) scale was developed. EFA, CFA, and reliability analyses based on a large-scale cross-sectional survey dataset (N = 555) allowed the identification of two factors: collaboration (10 items, α = 0.951) and shared responsibility (5 items, α = 0.879). Collaboration is defined as the joint interaction in the group in all activities that are needed to perform a shared task [4]. Shared responsibility means that colleagues create a common sense of responsibility for all students' learning [39]. The two-factor structure does fully align with the theoretically assumed two-dimensional structure [22]. The CSTT scale makes it, therefore, possible to assess collaboration and shared responsibility as two important dimensions of the team teaching practice. The development of the CSTT scale represents an advancement in the ability to assess and understand the subtleties of the team teaching practice. This scale serves as a specific tool to systematically assess the multifaceted aspects of collaboration and shared responsibility, two crucial dimensions that define the effectiveness of team teaching. In short, the CSTT scale serves as a lens through which it is possible to identify strengths as well as areas for improvement within the practice of team teaching. Its development enriches the toolkit available to both researchers and teachers. As a result, this study fills a gap in research and also enables teachers to develop more effective practices of team teaching.

Next, the first empirical insight into the practice of team teaching was provided. The results show that teachers with team teaching experience report a high degree of collaboration, and a high degree of shared responsibility. This means that teachers can count on each other for questions and concerns and give each other emotional and professional support. They mutually trust and respect each other, are open to reflection, and give each other feedback. It also implies that teachers are both responsible for the course or courses, and for their students' learning outcomes, well-being, and motivation. Previous research agrees that collaboration and shared responsibility can have a major impact on both teachers and students. For instance, the review study of Vangrieken, Dochy [4] shows that although achieving teacher collaboration proves challenging, it has many benefits for teachers and students, but also for the school. A recent study by Berry [79] indicates that a shared sense of responsibility for the education of students with disabilities can have positive effects on both teachers and students. It is particularly encouraging to note that teachers report high levels of collaboration and shared responsibility, since these are considered in the research literature as two important dimensions of the practice of team teaching [22]. Although most teachers report a high score for both dimensions of the practice of team teaching, there are also team teachers who report a lower score for a particular dimension, or even for both dimensions. The lower level of collaboration and shared responsibility could be explained by the team teaching model used. The models of team teaching represent the ways in which team teaching is established in the classroom (e.g., observation model,

parallel model, teaming model). For instance, the observation model would imply a lower level of collaboration and shared responsibility, compared to the teaming model [14]. In the observation model, one teacher observes while the other teacher teaches the course [13]. The focus of the observation is on the students. Further research measuring the relationship between the two could address this. The question could be raised whether it is necessary for both dimensions to score high in order to speak of quality team teaching. In our view, a lower score for one or both dimensions is not necessarily a negative sign. In this respect, it is important to emphasise that this measurement instrument is not normative, but rather seeks to reflect important dimensions of team teaching, without being all-encompassing.

Furthermore, tests for measurement invariance are reported on the two factors in the CSTT scale, providing support for configural, metric, scalar, and strict invariance by length of (a) teaching experience, (b) education type, and (c) frequency of team teaching. This means that teachers across these groups interpret the developed measurement instrument in a consistent manner. Therefore, it can be stated that the CSTT scale is a solid and robust instrument to be used with both experienced and less experienced teachers, with teachers from pre-primary, primary, secondary, and adult education, and with team teachers with both a low and a high frequency of team teaching. The CSTT scale is a 15-item scale, including 10 items to measure collaboration and five items for shared responsibility. It can be stated that its application is simple and fast, and it can be useful as a diagnostic measure, allowing the assessment of teachers' practice of team teaching. The CSTT scale has important implications for planning teaching and learning activities that contribute to improving the practice of teaching with respect to team teaching. For example, a team teaching team could use this scale as a tool to talk about their collaboration and shared responsibility as a team. If one or more teachers report, for instance, low(er) scores on collaboration compared with others, this may indicate a need to talk about it. To go deeper into conversation, even items can be discussed more concretely. A fairly low score on the item about discussing experiences openly could spark a conversation.

As measurement invariance of the CSTT scale is established across (a) teaching experience, (b) education type, and (c) frequency of team teaching, differences between these groups could be examined. Results indicate that there are no significant differences between the groups based on (a) teaching experience and (b) education type for both collaboration and shared responsibility. There are, however, significant differences between groups in terms of the (c) frequency of team teaching. Teachers who team teach less than once a week experience less collaboration and shared responsibility with their team teaching colleague(s), compared with teachers who team teach more than once a week. This finding suggests that teachers who frequently engage in team teaching experience more collaboration and shared responsibility.

Having conducted one of the first large-scale quantitative survey studies on team teaching, this study presents an instrument (i.e., the CSTT scale) to measure two important dimensions of the practice of team teaching. The development of this instrument is an important contribution to the field as it makes all kinds of new avenues of research possible to further investigate the practice of team teaching. Further research can, for example, investigate the relationship between the practice of team teaching and teachers' effective teaching behaviour. The first empirical insights show that team teachers experience a high degree of collaboration and shared responsibility. Additionally, the frequency of team teaching influences these dimensions of the practice of team teaching.

This study is not without limitations. First, although there are good theoretical reasons to believe that the practice of team teaching can be further conceptualised as collaboration and shared responsibility [22], other conceptualisations are also possible. For example, further research could take other dimensions such as team similarity, team efficacy, and team potency into account, as team teaching is a complex concept.

Second, it is necessary to be aware that the outcomes of the CSTT scale remain self-reported data. This implies that teachers' answers may have been influenced by social desirability, as is a risk with any form of subjective data collection [80]. However, through-

out the process of survey development and administration, several steps were taken to reduce social desirability bias. This included an expert review and a pilot study. Future research should combine this data with other data collection methods, such as observation or interview data. Results from other data collection methods can verify the validity of the CSTT scale.

Third, although the sample met all criteria required to develop the questionnaire, it solely consists of Flemish schools. This limits our claims to the generalisability of the questionnaire and the results to other contexts. Therefore, future research is encouraged to translate, adapt, and validate the CSTT scale in other educational settings. Moreover, the translation of the CSTT scale into different languages and its validation in different contexts will offer opportunities for additional and comparative research on the practice of team teaching in other regions and contexts. To facilitate this, the original Dutch version and an English translation are included as an Appendix. Authors should discuss the results and how they can be interpreted from the perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted. To adapt and use this scale in different cultures, contexts, and countries, researchers are advised to adopt a systematic process that takes into account both linguistic and cultural nuances. This procedure requires an extensive reiteration of the previously completed steps of this study, carefully considering linguistic and cultural nuances. Initially, the scale should be translated so that the essence of its constituent items is preserved. Next, pilot and/or expert testing is crucial to uncover possible language or comprehension problems. Finally, psychometric assessments must be conducted to determine the reliability and validity of the adapted scale.

7. Conclusions

This manuscript reports on the development of the Collaboration and Shared Responsibility in Team Teaching (CSTT) scale. The CSTT scale is an instrument to measure collaboration and shared responsibility, as two important dimensions of the team teaching practice. The first empirical evidence shows overall high scores on both dimensions. Further results indicate that there are no significant differences between the groups based on (a) teaching experience and (b) education type for both collaboration and shared responsibility. There are, however, significant differences between groups in terms of the (c) frequency of team teaching. In sum, the CSTT scale is a solid and robust instrument, which can be useful as a diagnostic measure to assess teachers' team teaching practice.

Appendix A

Table A1. Descriptive Statistics of the Items (n = 555).

	Item	Mean (SD)	Skewness	Kurtosis	Correlation																		
					C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	CS1	CS2	CS3	CS4	CS5	CS6	CS7	CS8
	During team teaching classes, my team teaching colleague(s) and I . . .																						
C1	can rely on each other for questions and concerns.	3.63 (0.68)	−2.20	5.63	1.00																		
C2	complement each other very well.	3.45 (0.75)	−1.44	1.96	0.70	1.00																	
C3	give each other emotional support.	3.41 (0.79)	−1.27	1.24	0.65	0.65	1.00																
C4	give each other professional support.	3.47 (0.72)	−1.42	2.03	0.70	0.72	0.66	1.00															
C5	mutually trust each other.	3.52 (0.72)	−1.47	1.71	0.71	0.72	0.69	0.73	1.00													
C6	respect each other.	3.70 (0.57)	−1.90	3.35	0.65	0.60	0.60	0.65	0.74	1.00												
C7	support each other.	3.58 (0.69)	−1.81	3.42	0.69	0.72	0.65	0.71	0.73	0.77	1.00											
C8	openly discuss experiences.	3.49 (0.75)	−1.53	2.15	0.69	0.66	0.62	0.69	0.71	0.64	0.70	1.00										
C9	are open to reflection.	3.48 (0.73)	−1.49	2.44	0.64	0.67	0.63	0.71	0.66	0.64	0.72	0.76	1.00									
C10	provide each other feedback.	3.22 (0.86)	−0.93	0.29	0.53	0.58	0.56	0.64	0.55	0.50	0.60	0.63	0.74	1.00								
C11	avoid competition.	3.68 (0.66)	−2.46	7.27	0.51	0.48	0.44	0.56	0.52	0.55	0.55	0.53	0.62	0.49	1.00							

Table A1. *Cont.*

	Item	Mean (SD)	Skewness	Kurtosis	Correlation																		
	During team teaching classes, my team teaching colleague(s) and I are both responsible for . . .																							
SR1	designing and preparing the lesson.	2.67 (1.32)	−0.68	−0.78	0.33	0.38	0.29	0.37	0.28	0.26	0.34	0.29	0.33	0.39	0.22	1.00							
SR2	classroom management.	3.19 (1.07)	−1.33	1.01	0.40	0.40	0.37	0.44	0.37	0.38	0.43	0.38	0.41	0.34	0.27	0.45	1.00						
SR3	student evaluation.	3.00 (1.11)	−1.00	0.19	0.28	0.32	0.30	0.33	0.25	0.24	0.27	0.31	0.31	0.37	0.17	0.57	0.46	1.00					
SR4	teaching and supporting the class assignments.	3.26 (0.96)	−1.46	1.84	0.36	0.37	0.25	0.40	0.29	0.27	0.34	0.29	0.33	0.35	0.23	0.58	0.54	0.55	1.00				
SR5	evaluating the lesson.	3.07 (1.01)	−1.00	0.44	0.37	0.46	0.36	0.47	0.34	0.32	0.40	0.39	0.48	0.49	0.31	0.65	0.46	0.56	0.62	1.00			
SR6	students' learning outcomes.	3.23 (0.96)	−1.39	1.70	0.36	0.35	0.34	0.37	0.30	0.29	0.29	0.33	0.34	0.38	0.18	0.53	0.46	0.69	0.54	0.64	1.00		
SR7	students' well-being.	3.53 (0.69)	−1.84	4.34	0.45	0.45	0.42	0.44	0.36	0.37	0.43	0.40	0.43	0.40	0.29	0.41	0.51	0.45	0.43	0.53	0.56	1.00	
SR8	students' motivation.	3.53 (0.93)	−1.73	3.98	0.45	0.49	0.42	0.48	0.43	0.42	0.46	0.44	0.48	0.45	0.31	0.43	0.55	0.47	0.50	0.56	0.59	0.86	1.00

Appendix B

Table A2. Results of Tests of Measurement Invariance by (a) Teaching Experience, (b) Education Type, and (c) Frequency of Team Teaching.

Multiple Group CFA	Overall Results						Model Difference Results				
	X^2 (df)	*p*	CFI	TLI	RMSEA	SRMR	Comparison	ΔCFI	ΔTLI	ΔRMSEA	ΔSRMR
Teaching experience											
Model 1 (configural invariance)	222.51 (178)	0.013	0.971	0.965	0.030	0.040					
Model 2 (metric invariance)	219.80 (191)	0.075	0.981	0.979	0.023	0.048	Model 2 vs. Model 1	0.010	0.014	0.007	0.007
Model 3 (scalar invariance)	233.37 (204)	0.077	0.981	0.980	0.023	0.048	Model 3 vs. Model 2	0.000	0.001	0.001	0.001
Model 4 (strict invariance)	245.49 (219)	0.106	0.983	0.983	0.021	0.051	Model 4 vs. Model 3	0.002	0.003	0.002	0.003
Education type											
Model 1 (configural invariance)	397.42 (356)	0.064	0.970	0.964	0.029	0.054					
Model 2 (metric invariance)	437.88 (395)	0.067	0.969	0.967	0.028	0.077	Model 2 vs. Model 1	0.001	0.002	0.001	0.024
Model 3 (scalar invariance)	495.08 (434)	0.022	0.955	0.957	0.032	0.081	Model 3 vs. Model 2	0.013	0.010	0.004	0.003
Model 4 (strict invariance)	550.23 (479)	0.013	0.948	0.954	0.033	0.094	Model 4 vs. Model 3	0.007	0.002	0.001	0.014
Frequency of team teaching											
Model 1 (configural invariance)	239.64 (178)	0.001	0.956	0.949	0.035	0.042					
Model 2 (metric invariance)	246.31 (191)	0.004	0.961	0.957	0.032	0.057	Model 2 vs. Model 1	0.004	0.008	0.003	0.015
Model 3 (scalar invariance)	269.76 (204)	0.001	0.954	0.952	0.034	0.059	Model 3 vs. Model 2	0.007	0.005	0.002	0.002
Model 4 (strict invariance)	294.01 (219)	0.001	0.947	0.949	0.035	0.066	Model 4 vs. Model 3	0.007	0.003	0.001	0.007

Appendix C

Table A3. The CSTT scale (English version).

	Item	Scale
	During team teaching classes, my team teaching colleague(s) and I . . .	
C1	can rely on each other for questions and concerns.	Collaboration
C2	complement each other very well.	Collaboration
C3	give each other emotional support.	Collaboration
C4	give each other professional support.	Collaboration
C5	mutually trust each other.	Collaboration
C6	respect each other.	Collaboration
C7	support each other.	Collaboration
C8	openly discuss experiences.	Collaboration
C9	are open to reflection.	Collaboration
C11	avoid competition.	Collaboration
	During team teaching classes, my team teaching colleague(s) and I are both responsible for . . .	
SR1	the design and preparation of the lesson.	Shared responsibility
SR3	student evaluation.	Shared responsibility
SR4	teaching and supporting the class assignments.	Shared responsibility
SR5	the evaluation of the lesson.	Shared responsibility
SR6	students' learning outcomes.	Shared responsibility

Appendix D

Table A4. De CSTT schaal (Nederlandstalige versie).

	Item	Schaal
	Tijdens teamteachingslessen, kunnen mijn teamteachingscollega('s) en ik . . .	
C1	bij elkaar terecht kunnen met vragen en bezorgdheden.	Samenwerking
C2	elkaar goed aanvullen.	Samenwerking
C3	elkaar emotionele steun geven.	Samenwerking
C4	elkaar professionele steun geven.	Samenwerking
C5	elkaar wederzijds vertrouwen.	Samenwerking
C6	elkaar respecteren.	Samenwerking
C7	elkaar ondersteunen.	Samenwerking
C8	ervaringen in alle openheid delen.	Samenwerking
C9	openstaan voor reflectie.	Samenwerking
C11	vermijden van competitie	Samenwerking
	Tijdens teamteachingslessen, zijn mijn teamteachingscollega('s) en ik samen verantwoordelijk voor . . .	
SR1	het ontwerpen en voorbereiden van de les.	Gedeelde verantwoordelijkheid
SR3	de evaluatie van de leerlingen.	Gedeelde verantwoordelijkheid
SR4	het geven en ondersteunen van de lesopdrachten.	Gedeelde verantwoordelijkheid
SR5	het reflecteren over de les.	Gedeelde verantwoordelijkheid
SR6	de leeruitkomsten van de leerlingen.	Gedeelde verantwoordelijkheid

References

1. Jang, J.; Yoo, H.; Rubadeau, K. How teacher collaboration profiles connect to literacy instructional practices: Evidence from PISA 2018 outcomes for Korea. *Int. J. Educ. Res.* **2022**, *114*, 102010. [CrossRef]
2. Vanblaere, B.; Devos, G. Exploring the link between experienced teachers' learning outcomes and individual and professional learning community characteristics. *Sch. Eff. Sch. Improv.* **2015**, *27*, 205–227. [CrossRef]
3. Deneire, A.; Vanhoof, J.; Faddar, J.; Gijbels, D.; Van Petegem, P. Characteristics of appraisal systems that promote job satisfaction of teachers. *Educ. Res. Perspect.* **2014**, *41*, 94–114.
4. Vangrieken, K.; Dochy, F.; Raes, E.; Kyndt, E. Teacher collaboration: A systematic review. *Educ. Res. Rev.* **2015**, *15*, 17–40. [CrossRef]

5. Ostovar-Nameghi, S.A.; Sheikhahmadi, M. From Teacher Isolation to Teacher Collaboration: Theoretical Perspectives and Empirical Findings. *Engl. Lang. Teach.* **2016**, *9*, 197. [CrossRef]
6. Johnson, B. Teacher collaboration: Good for some, not so good for others. *Educ. Stud.* **2003**, *29*, 337–350. [CrossRef]
7. Sehgal, P.; Nambudiri, R.; Mishra, S.K. Teacher effectiveness through self-efficacy, collaboration and principal leadership. *Int. J. Educ. Manag.* **2017**, *31*, 505–517. [CrossRef]
8. Reeves, P.M.; Pun, W.H.; Chung, K.S. Influence of teacher collaboration on job satisfaction and student achievement. *Teach. Teach. Educ.* **2017**, *67*, 227–236. [CrossRef]
9. van Garderen, D.; Stormont, M.; Goel, N. Collaboration between general and special educators and student outcomes: A need for more research. *Psychol. Sch.* **2012**, *49*, 483–497. [CrossRef]
10. Goddard, Y.L.; Goddard, R.D.; Tschannen-Moran, M. A Theoretical and Empirical Investigation of Teacher Collaboration for School Improvement and Student Achievement in Public Elementary Schools. *Teach. Coll. Rec. Voice Scholarsh. Educ.* **2007**, *109*, 877–896. [CrossRef]
11. Krammer, M.; Rossmann, P.; Gastager, A.; Gasteiger-Klicpera, B. Ways of composing teaching teams and their impact on teachers' perceptions about collaboration. *Eur. J. Teach. Educ.* **2018**, *41*, 463–478. [CrossRef]
12. Solis, M.; Vaughn, S.; Swanson, E.; Mcculley, L. Collaborative models of instruction: The empirical foundations of inclusion and co-teaching. *Psychol. Sch.* **2012**, *49*, 498–510. [CrossRef]
13. Honigsfeld, A.; Dove, M.G. Preparing Teachers for Co-Teaching and Collaboration. In *The Handbook of TESOL in K-12*; de Oliveira, L., Ed.; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2019; pp. 405–421. [CrossRef]
14. Baeten, M.; Simons, M. Student teachers' team teaching: Models, effects, and conditions for implementation. *Teach. Teach. Educ.* **2014**, *41*, 92–110. [CrossRef]
15. Jang, S.-J. Research on the effects of team teaching upon two secondary school teachers. *Educ. Res.* **2007**, *48*, 177–194. [CrossRef]
16. Murchú, F.; Conway, P. (Re)positioning team teaching: The visibility and viability of learning in classrooms. *Educ. Res. Perspect.* **2017**, *44*, 43–69.
17. Simons, M.; Baeten, M.; Vanhees, C. Team Teaching During Field Experiences in Teacher Education: Investigating Student Teachers' Experiences with Parallel and Sequential Teaching. *J. Teach. Educ.* **2020**, *71*, 24–40. [CrossRef]
18. Lieberman, A.; Miller, L. *Teachers in Professional Communities: Improving Teaching and Learning*; Teachers College Press: New York, NY, USA, 2008.
19. Rickard, A.; Walsh, T. Policy, practice and process in team teaching: A pilot project with co-operating teachers and student teachers on school placement. *Ir. Educ. Stud.* **2019**, *38*, 309–326. [CrossRef]
20. Veteska, J.; Kursch, M.; Svobodova, Z.; Tureckiova, M.; Paulovcakova, L. Longitudinal co-teaching projects—Scoping review. In *Cognition and Exploratory Learning in the Digital Age*; Springer: Berlin/Heidelberg, Germany, 2020.
21. Sweigart, C.A.; Landrum, T.J. The Impact of Number of Adults on Instruction: Implications for Co-Teaching. *Prev. Sch. Fail. Altern. Educ. Child. Youth* **2015**, *59*, 22–29. [CrossRef]
22. Simons, M.; Coetzee, S.; Baeten, M.; Schmulian, A. Measuring learners' perceptions on a team taught learning environment Development and validation of the LTTPQ. *Learn. Environ. Res.* **2019**, *23*, 45–58. [CrossRef]
23. Meirsschaut, M.; Ruys, I. *Team Teaching: Wat, Waarom, Hoe en Met Welke Resultaten? Een Verkenning van de Literatuur*; SONO: Ghent, Belgium, 2017.
24. Welch, M.; Brownell, K.; Sheridan, S. What's the score and game plan on teaming in schools. *Remedial Spec. Educ.* **1999**, *20*, 36–49. [CrossRef]
25. Thousand, J.S.; Villa, R.A.; Nevin, A.I. The Many Faces of Collaborative Planning and Teaching. *Theory Pract.* **2006**, *45*, 239–248. [CrossRef]
26. Fuller, R.; Bail, J. Team Teaching in the Online Graduate Environment: Collaborative Instruction. *Int. J. Inf. Commun. Technol. Educ.* **2011**, *7*, 72–83. [CrossRef]
27. Walsh, T. 'Promoted widely but not valued': Teachers' perceptions of team teaching as a form of professional development in post-primary schools in Ireland. *Prof. Dev. Educ.* **2020**, *48*, 688–704. [CrossRef]
28. Carless, D.R. Good practices in team teaching in Japan, South Korea and Hong Kong. *System* **2006**, *34*, 341–351. [CrossRef]
29. York-Barr, J.; Ghere, G.; Sommerness, J. Collaborative Teaching to Increase ELL Student Learning: A Three-Year Urban Elementary Case Study. *J. Educ. Stud. Placed Risk JESPAR* **2007**, *12*, 301–335. [CrossRef]
30. Goodnough, K.; Osmond, P.; Dibbon, D.; Glassman, M.; Stevens, K. Exploring a triad model of student teaching: Pre-service teacher and cooperating teacher perceptions. *Teach. Teach. Educ.* **2009**, *25*, 285–296. [CrossRef]
31. Bullough, R.V.; Young, J.; Birrell, J.R.; Clark, D.C.; Egan, M.W.; Erickson, L.; Frankovich, M.; Brunetti, J.; Welling, M. Teaching with a peer: A comparison of two models of student teaching. *Teach. Teach. Educ.* **2003**, *19*, 57–73. [CrossRef]
32. Gladman, A. Team Teaching Is Not Just for Teachers! Student Perspectives on the Collaborative Classroom. *TESOL J.* **2015**, *6*, 130–148. [CrossRef]
33. De Backer, L.; Simons, M.; Schelfhout, W.; Vandervieren, E. Let's team up Measuring student teachers' perceptions of team teaching experiences. In *Teacher Education: New Perspectives/Kayapinar*; IntechOpen: London, UK, 2021.
34. Higgins, L.M.; Litzenberg, K.K. Transferring Experience Through Team Teaching: The Chance of a Lifetime. *Coll. Teach.* **2015**, *63*, 105–111. [CrossRef]

35. Friend, M.; Cook, L. *Interactions: Collaboration Skills for School Professionals*, 4th ed.; Allyn and Bacon: Boston, MA, USA, 2003; p. 186.
36. Kelchtermans, G. Teacher collaboration and collegiality as workplace conditions: A review. *Z. Pädagogik* **2006**, *52*, 220–237.
37. Wullschleger, A.; Vörös, A.; Rechsteiner, B.; Rickenbacher, A.; Merki, K.M. Improving teaching, teamwork, and school organization: Collaboration networks in school teams. *Teach. Teach. Educ.* **2023**, *121*, 103909. [CrossRef]
38. Valckx, J.; Devos, G.; Vanderlinde, R. Exploring the relationship between professional learning community characteristics in departments, teachers' professional development, and leadership. *Pedagog. Stud.* **2018**, *95*, 34–55.
39. Stoll, L.; Bolam, R.; McMahon, A.; Wallace, M.; Thomas, S. Professional Learning Communities: A Review of the Literature. *J. Educ. Chang.* **2006**, *7*, 221–258. [CrossRef]
40. Sleegers, P.; den Brok, P.; Verbiest, E.; Moolenaar, N.M.; Daly, A.J. Toward Conceptual Clarity. *Elem. Sch. J.* **2013**, *114*, 118–137. [CrossRef]
41. Griffin, P.; Robertson, P. Professional learning teams and decision making. In *Assessment for Teaching*; Griffin, P., Ed.; Cambridge University Press: New York, NY, USA, 2014; pp. 13–25.
42. Bolam, R.; McMahon, A.; Stoll, L.; Thomas, S.; Wallace, M.; Greenwood, A.; Hawkey, K.; Ingram, M.; Atkinson, A.; Smith, M. *Creating and Sustaining Effective Professional Learning Communities*; University of Bristol: Bristol, UK, 2005; p. 222.
43. Gehlbach, H.; Brinkworth, M.E. Measure Twice, Cut down Error: A Process for Enhancing the Validity of Survey Scales. *Rev. Gen. Psychol.* **2011**, *15*, 380–387. [CrossRef]
44. Vangrieken, K.; Grosemans, I.; Dochy, F.; Kyndt, E. Teacher autonomy and collaboration: A paradox? Conceptualising and measuring teachers' autonomy and collaborative attitude. *Teach. Teach. Educ.* **2017**, *67*, 302–315. [CrossRef]
45. American Educational Research Association; American Psychological Association; National Council on Measurement in Education. *Standards for Educational and Psychological Testing*; American Educational Research Association: Washington, DC, USA, 2014.
46. Kline, R.B. *Principles and Practice of Structural Equation Modeling*; Little, T., Ed.; The Guilford Press: New York, NY, USA, 2016.
47. Tabachnick, B.; Fidell, L. *A Practical Approach to Using Multivariate Analyses*; Pearson Education: Boston, MA, USA, 2013; p. 983.
48. Kaiser, H.F. An index of factorial simplicity. *Psychometrika* **1974**, *39*, 31–36. [CrossRef]
49. Bartlett, M. The effect of standardization on a X2 approximation in factor analysis. *Biometrika* **1951**, *38*, 337–344.
50. Field, A.; Gaitanou, C. *Discovering Statistics Using SPSS*; Sage Publications: London, UK, 2009; Volume 3. [CrossRef]
51. Osborne, J. *Best Practices in Exploratory Factor Analysis*; CreateSpace Independent Publishing Platform: Louisville, KY, USA, 2014.
52. Revelle, W. *Psych: Procedures for Psychological, Psychometric, and Personality Research*; Northwestern University: Evanston, IL, USA, 2022.
53. O'Connor, B.P. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav. Res. Methods, Instrum. Comput.* **2000**, *32*, 396–402. [CrossRef]
54. Kaiser, H.F. The Application of Electronic Computers to Factor Analysis. *Educ. Psychol. Meas.* **1960**, *20*, 141–151. [CrossRef]
55. Cattell, R.B. The Scree Test for The Number of Factors. *Multivar. Behav. Res.* **1966**, *1*, 245–276. [CrossRef]
56. Hayton, J.C.; Allen, D.G.; Scarpello, V. Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organ. Res. Methods* **2004**, *7*, 191–205. [CrossRef]
57. Horn, J.L. A rationale and test for the number of factors in factor analysis. *Psychometrika* **1965**, *30*, 179–185. [CrossRef] [PubMed]
58. Velicer, W.F. Determining the number of components from the matrix of partial correlations. *Psychometrika* **1976**, *41*, 321–327. [CrossRef]
59. Velicer, W.; Eaton, C.; Fava, J. Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In *Problems and Solutions in Human Assessment*; Jackson, D.N., Goffin, R., Helmes, E., Eds.; Academic/Plenum Publishers, Kluwer: Alphen aan den Rijn, Netherlands, 2000; pp. 41–71.
60. Hair, J.F. *Multivariate Data Analysis*; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2010; Volume 7.
61. Rosseel, Y. lavaan: An R package for structural equation modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. [CrossRef]
62. Bentler, P.M. Comparative Fit Indexes in Structural Models. *Psychol. Bull.* **1990**, *107*, 238–246. [CrossRef]
63. Tucker, L.R.; Lewis, C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* **1973**, *38*, 1–10. [CrossRef]
64. Steiger, J.H. Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivar. Behav. Res.* **1990**, *25*, 173–180. [CrossRef]
65. Jöreskog, K.; Sörbom, D. *LISREL V: Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Methods*; University of Uppsala, Department of Statistics: Uppsala, Sweden, 1981.
66. Hooper, D.; Coughlan, J.; Mullen, M. Structural equation modelling: Guidelines for determining model fit. *Electron. J. Bus. Res. Methods* **2008**, *6*, 53–60.
67. Schermelleh-Engel, K.; Moosbrugger, H.; Muller, H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res.* **2003**, *8*, 23–74.
68. Hu, L.T.; Bentler, P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* **1999**, *6*, 1–55. [CrossRef]
69. De Neve, D.; Devos, G. How do professional learning communities aid and hamper professional learning of beginning teachers related to differentiated instruction? *Teach. Teach.* **2016**, *23*, 262–283. [CrossRef]

70. Thomas, L.; Rienties, B.; Tuytens, M.; Devos, G.; Kelchtermans, G.; Vanderlinde, R. Unpacking the dynamics of collegial networks in relation to beginning teachers' job attitudes. *Res. Pap. Educ.* **2021**, *36*, 611–636. [CrossRef]

71. Byrne, B.; Shavelson, R.J.; Muthén, B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychol. Bull.* **1989**, *105*, 456–466. [CrossRef]

72. Van de Schoot, R.; Lugtig, R.; Hox, J. A checklist for testing measurement invariance. *Eur. J. Dev. Psychol.* **2012**, *9*, 486–492. [CrossRef]

73. Meredith, W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika* **1993**, *58*, 525–543. [CrossRef]

74. Svetina, D.; Rutkowski, L.; Rutkowski, D. Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines: An Illustration Using Mplus and the lavaan/semTools Packages. *Struct. Equ. Model. A Multidiscip. J.* **2020**, *27*, 111–130. [CrossRef]

75. Cheung, G.W.; Rensvold, R.B. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Struct. Equ. Model. A Multidiscip. J.* **2002**, *9*, 233–255. [CrossRef]

76. Chen, F.F. Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Struct. Equ. Model. A Multidiscip. J.* **2007**, *14*, 464–504. [CrossRef]

77. Kassambara, A. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*, 2021. Available online: https://rpkgs.datanovia.com/rstatix/ (accessed on 23 July 2023).

78. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Erlbaum: Hillsdale, MI, USA, 1988.

79. Berry, A.B. Understanding Shared Responsibility Between Special and General Education Teachers in the Rural Classroom. *Rural. Spéc. Educ. Q.* **2021**, *40*, 95–105. [CrossRef]

80. Desimone, L.M. Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educ. Res.* **2009**, *38*, 181–199. [CrossRef]