CHAPTER THREE

METHODOLOGY OF THE STUDY

3.1 Introduction

In the previous section, the review of literature related to this study was presented. Review of related literature enables us to learn how other researchers construct their studies and provides us with useful examples and models for our own research (Creswell, 2015). Research methodology is one of these examples for researchers. The research methodology is a set of rules and procedures that guide researchers in the development of appropriate tools, the systematic collection of data from appropriate sources, and the measurement of variables. This will provide a framework for deriving conclusions (Miller & Brewer, 2003). However, prior to these, it is very much required to introspect on the purpose of the research and in which context the research was carried out.

3.2 Context and purpose of the study:

This research was indeed executed to examine the current status of students' engagement in learning, students' perception of teachers' engagement in the classroom teaching process and their influence on achievement of secondary school students of West Bengal, India. This research addressed only the government secondary level schools under West Bengal Board of Secondary Education (WBBSE) having rural and urban compositions of schools. A uniform school curriculum as well as uniform evaluation procedures was followed in all the schools. More specifically, the annual examination of class-X (i.e. Madhyamik Pariksha) is the first public examination in West Bengal that is conducted on the basis of identical sets of question papers following a uniform syllabus on a particular set of dates across the entire state.

The study attempted to quantify the relationship between the process of students' engagement in classroom teaching-learning transactions and the outcome of the classroom teaching-learning process in terms of academic achievement in reference to the contextual factor students' perception of teacher engagement. None of the variables was neither manipulated nor controlled, either directly or indirectly. However, students'

academic achievement was confined to the percentage of marks secured by the students in their final annual examination (Madhyamik Pariksha) of 10th grade.

Figure 3.1

Location map for the selected districts of West Bengal for data collection

In sum, the review of the literature illustrated a need to develop a greater understanding in how students' perception of teacher engagement in teaching and student engagement in learning affect academic success for students. Based on this gap in the research, the intent of this investigation was to investigate the associations among teacher engagement, student engagement and academic achievement of the students. This led to the following research questions:

1. Is there any gender gap in perceived teacher engagement? Who perceived teachers more engaged in teaching among boys and girls?
2. Is there any gender gap in student engagement? Who are at the risk of lower engagement among boys and girls?
3. Is there any gender gap in academic achievement? How students' academic achievement differ for the students across their gender?
4. Whether and how the gender gap in perceived teacher engagement dimension(s) may explain gender gap in student engagement?
5. Whether there are differential effects of students' perceived teacher engagement on boys' as compared to girls' engagement?
6. Whether gender gap in student engagement acts as an explaining mechanism in gender gap in academic achievement?
7. Which student engagement dimension(s) matter more for the boys than the girls in enhancing their academic achievement?
8. Whether students' engagement explains the association between teacher engagement and achievement of the students?

3.3 Roadmap of the Chapter:

This chapter is the most critical part as it provides the description of the methods and procedures followed in the present study. This section basically narrates 'why' and 'how' of the decisions made regarding sampling procedures used to draw a sample from the population, the procedures how the data were collected, research philosophy followed in the study, research design for this study, research approach adopted in this study, research method employed in the study and regarding data analysis strategies, the measurement of the latent variables associated with the study especially, the procedures of development

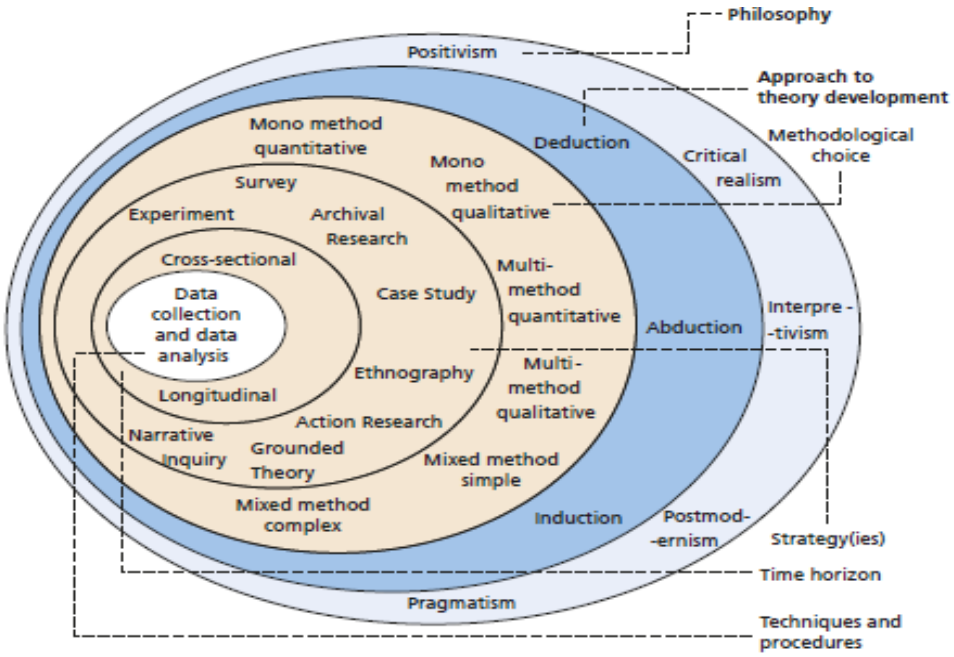
and validation of the research instruments and evaluation of their psychometric properties. In this regard, the 'research onion' framework (Saunders, Lewis, & Thornhill, 2019) was followed.

3.4 Philosophical and methodological choices: The 'Research Onion' framework

In this research, several crucial and systematic decisions were taken to answer the research questions. For this, 'Research onion' framework (see Figure 3.2) was followed (Saunders, Lewis, & Thornhill, 2019). It provides the scope to justify each of those choices and also enables to make aligned decisions. The 'research onion' framework explains how to make appropriate research choices to answer specific research questions. The framework consists six consecutive layers (see Figure 3.2): i) research philosophy, ii) approach to theory development, iii) methodological choice(s), iv) research strategies, v) time horizon, vi) techniques and procedures of data collection and data analysis. Importantly, each decision regarding the study must be synchronized with the purpose of research, research objectives and with the research questions.

Figure 3.4

The 'Research Onion' adapted from Saunders, Lewis, and Thornhill (2019, p. 130)



3.4.1 Research paradigm and Research philosophy

Research philosophy is a way of seeing the reality and truth (ontology), way of gaining knowledge or understanding of the reality or truth (epistemology), and a way how researcher can contribute to this research (axiology). Research philosophy should be informed by the research aims and research questions. The current study is associated with the research questions that seek for objective and quantifiable answers. The current study indeed attempted to quantify the interrelationships among students' perceptions of teacher engagement, their engagement and achievement in school. Thus, positivism appears to be more appropriate in this case. Positivism is associated with scientific methods and believes in a single truth or single reality. A researcher who follows positivism needs to be objective and unbiased. Positivism deals with theory or hypothesis that one can go and test with numbers. In Positivist paradigm of research participants' views, opinion and perceptions collected and quantified with the belief that the respondents' perceptions can tell a lot about how they think, behave and act in a particular situation.

3.4.2 Research approach

The present study followed a quantitative approach as it dealt with numerically quantifiable variables. Quantitative approach was employed as the variables involved in the study can be quantitatively measured and the objectives of the study are quantifiable (Creswell, 2015). Indeed, the researcher attempted to make meaning of the reality out of the numbers assigned by measuring attributes. The research questions were premised on the voids in the existing literature and this study actually tested theories based on the collected data. So, a deductive approach was followed in this research. Besides, Confirmatory research approach follows a deductive technique that tests hypotheses and confirms theories based on the collected quantitative data.

3.4.3 Methodological choices

This layer of 'Research onion' is about deciding how many data analysis methods the researcher will make use of. While selecting the research method one must become purposeful in answering the research questions in a meaningful and effective way. This study dealt with research questions associated with the extent or the degree of

relationship among the study variables which are quantifiable in nature. Thus, the researcher employed quantitative method in this study.

This study adopted the descriptive research approach because it provides the correct and appropriate explanation of the results (Creswell, 2015). It emphasizes what currently exists, such as realistic conditions, behaviors, situations, or phenomena. Since this study was concerned with the present status of perceived teacher engagement, student engagement and their relationships with academic achievement of students, it seemed the most appropriate to employ the descriptive method for this research. In this study, the researcher did not manipulate any variable rather recorded the perception, belief, and opinion of the students regarding their engagement in learning, how they perceived that teachers' are engaged in teaching in classroom settings and investigated how all these influence students' academic achievement. Responses of the students were collected in the natural settings without any intervention by the researcher.

The survey research worked well for the study as it provided the respondents' feelings about how teachers were engaged and at the same time how students were engaged in their learning. The quantification of the opinion enabled the researcher to answer the research questions by explaining the relationships among the study variables, quantitatively.

3.4.4 Research strategy

Research strategy deals with practical and tangible aspects of research decisions. This study followed descriptive research strategy where survey was undertaken on a sample that is expected to be representative of a larger population. Descriptive researches seek to describe attributes of a larger population of interest. This study employed survey research method, specifically a cross-sectional survey, in which "the researcher collected data at a single point in time" (Creswell, 2015, p. 377) and analyzed the data to describe recent opinions to statements, research questions, and hypotheses. To check the status of the phenomenon, a descriptive survey was done. Survey research assists in gathering information at a time, summarizing, presenting, and interpreting it in accordance with the objectives set to explore, describe, and explain respondents' behavior. Feedback on the research questions, demographic information, and academic achievement of students

were collected from survey participants and school administrations for the further analysis.

3.4.5 Time horizon

For this study, collection of data was carried out almost at one specific point of time to unpack/showcase description about how people feel about an issue at that point of time. Hence, a cross-sectional survey was found suitable for the research where the impact of time is not a variable under study.

3.4.6 Techniques and procedures:

This layer is concerned with the practical details of data collection, how data were collected and which steps were followed during the field work and how the collected data were analyzed. Consents were taken from both the head of the institutions and of the students for getting the responses. Then, students were provided with the research tools and were asked to follow the instructions before providing their responses. Thus, data collection was completed in paper-pencil mode in classroom context. Further, students were informed that their responses will not be evaluated or disclosed and will be used for academic purposes only. Details of data collection process have been discussed in section 3.5.3. Further, collected data were analyzed using SPSS which is a statistical program to do a bunch of inferential tests. For data analysis, independent samples t-test for equality of means was to test $H_0$1-3, mediation analysis was to test $H_0$4, $H_0$6, and $H_0$8 and moderation analysis was used to test $H_0$5 and $H_0$7.

3.5 Research design

To find the answers of the research questions, the current study employed correlational design which is basically a quantitative and non-experimental research design. The correlational design was found appropriate to describe, explore, and explain the associations among the criterion variables teacher engagement, student engagement and academic achievement. The setting of the study was the tenth graders' classrooms in government secondary schools in West Bengal under WBBSE. However, neither the variables nor the context of data collection was controlled by the researcher by any means while collecting responses from the students.

This study adopted a quantitative research design in that the researcher utilized measurement of the criterion variables to test the hypotheses by gathering data via surveys and analyzing it using statistical methods (Creswell, 2009). Another explanation is because this study involves a statistical, mathematical, or computational exploration of observable events in a methodical manner (Best & Kahn, 2016). This research design was suitable for the study since its objectives was to examine the impact of teacher engagement and student engagement on their achievement. In addition, it was non-experimental as the variables in the study, perceived teacher engagement and student engagement of tenth graders were studied in their natural settings, thus, those were not controlled nor influenced (Belli, 2008).

The present study attempted to investigate how students' perceived that how teachers are engaged in teaching, how the perception of teachers' engagement influenced students' engagement in their learning that further influenced their learning outcomes that may be reflected in their academic achievement. Hence, to address the research questions/objectives, the descriptive survey design was employed in the current study (see Figure 3.3). This design was found appropriate for the study because the primary focus of these studies remains on exploring existing facts and identifying relationships between different variables in a control-free environment. Furthermore, another focus of descriptive studies is fact generalization (Creswell, 2015) across large population.

Figure 3.3

Research design

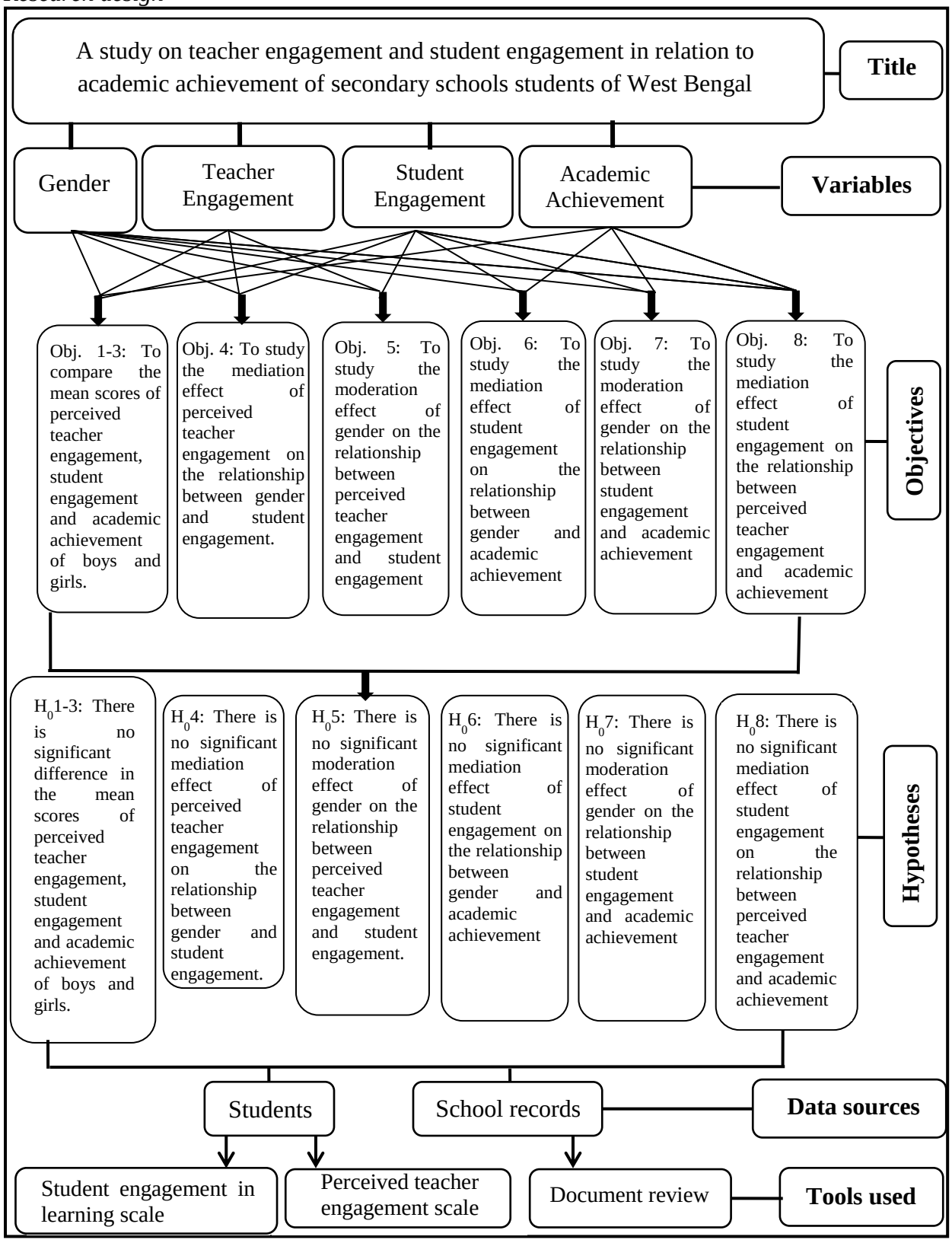
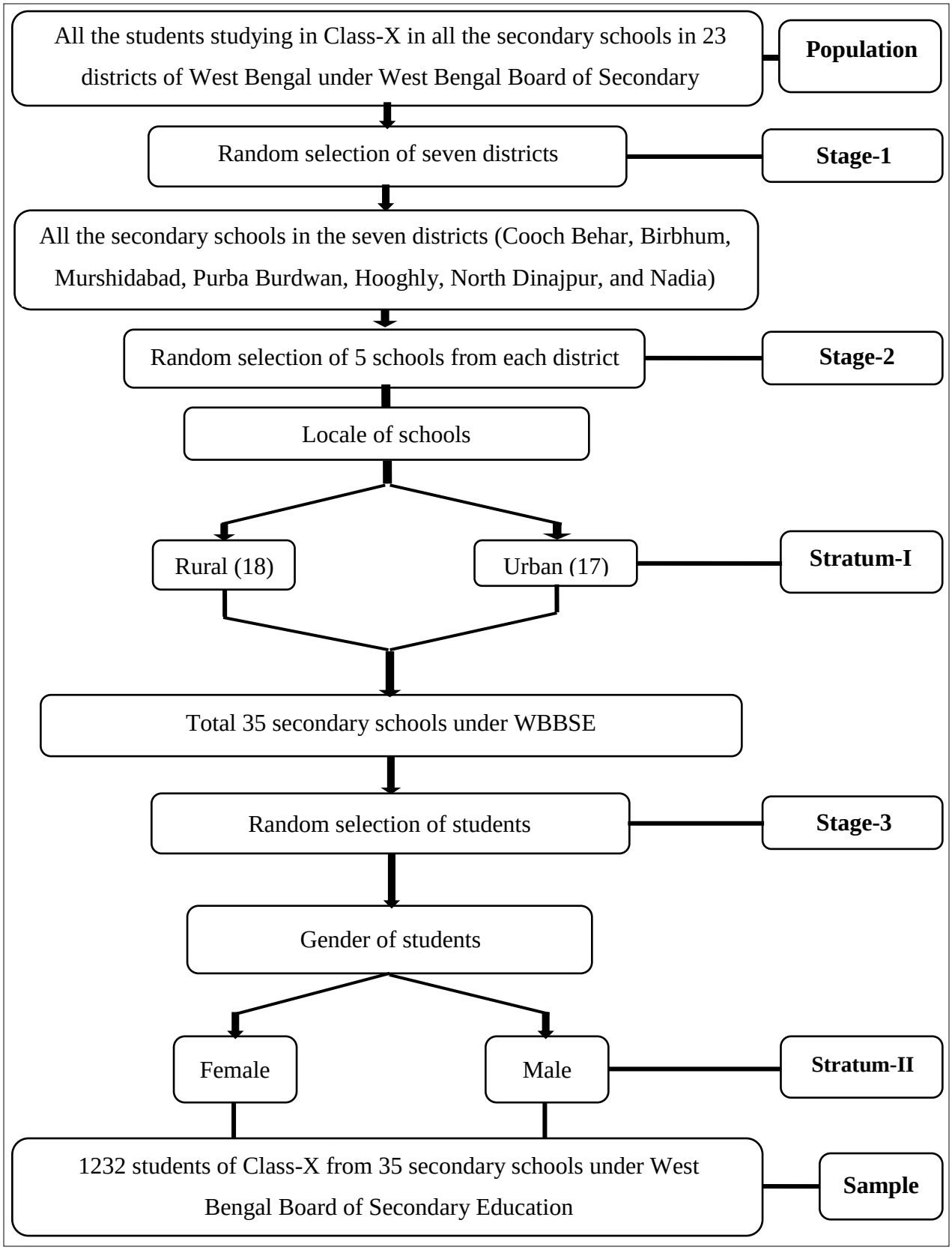


Figure 3.4

Sampling design of the study



3.5.1 Population, sampling procedures and sample of the study

The research was conducted in the context of secondary schools under West Bengal Board of Secondary Education (WBBSE). A uniform school curriculum as well as uniform evaluation procedures was followed in all the schools. Bengali was the medium of instruction and language of communication in all the selected schools. The sample was selected using multistage sampling procedures. In the first stage, 7 districts (viz. Cooch Behar, Birbhum, Murshidabad , Purba Bardwan, Hooghly, North Dinajpur, and Nadia) out of 23 districts in West Bengal was randomly selected using lottery method (see Figure 3.4). The location map has been shown in Figure 3.1 to specify the location where the study was conducted. In the second stage, 5 schools from each district were randomly selected from the lists of schools from the school directory available from http://wbbse.wb.gov.in. The schools were selected following three criteria: (1) geographical distribution of schools in West Bengal, and (2) locale of schools (rural and urban). In the next stage, 40 students were randomly selected from each school. However, 106 students left student engagement in learning scale (SELS) incomplete, 39 students filled demographic information incompletely, and 23 students did not respond to all items in perceived teacher engagement scale (PTES). Overall, 1232 students completed the research tools. In the present study, the participants were 1232 tenth-grade students with 611 female (49.594%) and 621 male (50.406%) students from 35 government secondary schools (Rural=18, Urban=17) across the seven districts of West Bengal, India (see Table 3.1).

Studies (e.g. Krejcie & Morgan, 1970) have recommended that a sample size of 384 sufficiently represents a large population of 10,000,000 respondents or more. However, the present study is a survey research that is characterized by generalizations of its outcomes over a large population (Creswell, 2015). Hence, a larger sample size is recommended (Best & Kahn, 2016). For this study the sample size was 1232.

3.5.2 Delimitations of the study

This research was delimited to the Bengali medium regular secondary level schools which are regulated by West Bengal Board of Secondary Education (WBBSE). Further, the study was also delimited only to the 10^{th} graders in those secondary schools.

Table 3.1

Demographic profile of the participants

	Sample (1232 participants)	
	N	%
Gender		
Female	611	49.594
Male	621	50.406
Locale of school		
Rural (18 schools)	617	50.081
Urban (17 schools)	615	49.919
Geographical location (Districts)		
Cooch Behar	167	13.555
Birbhum	175	14.205
Murshidabad	172	13.961
Purba Bardwan	187	15.179
Hooghly	180	14.610
North Dinajpur	183	14.854
Nadia	168	13.636

3.5.3 Measurements and Research instruments

In order to quantify the associations among the study variables and to answer the research questions formulated in this research, the study variables were measured quantitatively using the following research tools: PTES was to gauge students' perceived teacher engagement, SELS was used for measuring student engagement, and students' academic achievement were collected from the school records. The research instruments consisted of both positively and negatively worded items, hence reverse scoring was used before performing any statistical analysis. In all the instruments, the researcher added questions related to demographic factors like gender, caste, name of school, and locale of school.

A. Student engagement in Learning scale (SELS)

Student engagement was gauged using researcher's self-developed 23-item scale (11 negative items) consisting of three sub-scales: cognitive (7 items, 2 negatives), behavioral (8 items, 5 negatives), and emotional (8 items, 4 negatives) engagement rated on a 5-point Likert scale from 1= '*Strongly disagree'* to 5= '*Strongly agree'*. Some items are: 'Whenever I find something difficult, I give up right away.' (Cognitive engagement), 'I wait for the moment when class finishes.' (Behavioral engagement) and 'It is better to read at home than attending class.' (Emotional engagement). The Cronbach's Alpha (Cronbach. 1951) values for the engagement sub-scales were .90, .92 and .92 for cognitive, behavioral and emotional engagement, respectively, whereas, .88 for the overall scale.

B. Perceived teacher engagement scale (PTES)

Students' perceived teacher engagement was gauged using researcher's self-developed 22-item scale (11 negative items) consisting of three sub-scales: perceived cognitive-physical (7 items, 4 negatives), perceived socio-emotional (8 items, 3 negatives), and perceived pedagogical (7 items, 4 negatives) engagement rated on a 5-point Likert scale from 1= '*Strongly disagree'* to 5= '*Strongly agree'*. Some items are: 'Teacher puts all effort to make us understand lessons.' (Perceived Cognitive-physical engagement), 'Teacher follows whether I am doing my tasks in class.' (Perceived Socio-emotional engagement) and 'Teacher provides interesting activities in class.' (Perceived Pedagogical engagement). The Cronbach's Alpha values for the engagement sub-scales were .91, .93 and .91 for perceived cognitive-physical, perceived socio-emotional and perceived pedagogical engagement, respectively, whereas, .87 for the overall scale.

C. Academic achievement

The percentage of marks secured by the students in the final annual examination of the students was considered as the academic achievement of the students.

D. Gender

Students reported their gender in the specified columns in both of the above-mentioned tools. For the entire dataset, females were coded as 1 and males as 2.

3.5.4 Procedures of data collection

In the present study, the field data were collected from the tenth-grade students from the 31 selected secondary schools to test hypotheses 1 to 9. The following protocols were maintained during data collection. To get permissions for collecting data, the headmasters/principals of the selected schools were contacted and acquainted with the purposes of this study and were asked to provide their compliance and co-operation in the due course of action. Then, consents were taken from the students. Among them who wanted to participate voluntarily, twenty were selected at purely random basis and were requested to fill the research tools. They were asked to provide their responses about their engagement in the classroom learning. They were instructed to provide their responses to each item only after completing their demographic information. Further, to minimize social desirability bias, they were also instructed that their responses were neither going to be evaluated nor to be disclosed elsewhere. Additionally, the students were requested to reflect on their honest perceptions. However, students were not entertained with any kind of incentives for providing their responses. Further, to elevate the rate of honest responses reducing social desirability bias, students were also intimated that their responses will not be evaluated as right or wrong.

The participants were also informed that all the responses will be kept confidential. The tools were given only to the students who participated voluntarily. They were asked to complete their demographic information before going to the scale items. However, no incentives were provided for filling the scale. This formal investigation was executed in a classroom setting and the entire session was supervised by the investigators. Finally, a sample of 1232 participants was obtained who completed both scales. Out of the sample, 611 participants (49.594%) were females and the rest 621 participants (50.406%) were males (for details see Table 3.1).

The previous researches (NICHD Early Child Care Research Network, 2005) suggested that although classroom observation measures behavioral engagement however it is not an appropriate tool for assessing internal psychological experiences and emotional activities of the students. Again, teacher reports reflect cumulative ratings of engagement over the academic session that incorporates subjective perceptions of their students

(Mashburn, Hamre, Downer, & Pianta, 2006). Further, some researchers (Valiente, Lemery-Chalfant, Swanson, & Reiser, 2008) cautioned about over-reliance on teacher-reported data. Hence, only the student report of their engagement in Mathematics class was considered in the present study.

In summary, the researcher collected data from two sources: primary data from students using the questionnaires, and secondary data from school records in the form of students' annual marks.

3.5.5 Data analysis strategy

In this study, SPSS, Version 26.0 for the entire data analysis and AMOS 23.0 was used to perform Confirmatory factor analysis of the measurement models of student engagement and perceived teacher engagement. The significance level was set at α=0.05 for all the statistical comparisons.

To investigate whether girls and boys have different levels of engagement, perceived teacher engagement and their achievement, the statistical significance of the differences of the mean scores of the variables were examined using independent samples t-test for equality of means ($H_0$1-3). However, $H_0$5, and $H_0$7 were tested using moderation analysis, whereas, $H_0$4, $H_0$6, and $H_0$8 were examined with the help of mediation analysis. The mediation hypotheses ($H_0$4, $H_0$6, and $H_0$8) were tested following parallel mediation analysis using model 4 in PROCESS macro for SPSS (developed by Prof. Andrew F. Hayes; Hayes, 2013). PROCESS works on regression-based path-analytic framework that estimates the indirect effect with bias-corrected confidence intervals. To examine the statistical significance of the direct, indirect and total effects, Bias-corrected Bootstrapping resampling methods were used. The guidelines of Preacher et al. (2007) and Hayes (2013) were followed to perform the bootstrapping resampling procedures (on 10,000 Bootstrap sub-samples) that produced 95% bias-corrected confidence intervals. The total, direct and indirect effects were considered statistically significant at α=.05 when the corresponding bias-corrected confidence intervals of the effects did not include zero (Field, 2013). Bootstrapping method was intentionally selected as it is a nonparametric resampling procedure that does not violate assumptions of normality (Koopman et al., 2015). Nevertheless, the mediation with a non-significant direct effect

and significant indirect effect refers to 'full mediation' whereas, mediation with significant direct effect and significant indirect effects refers to 'Partial mediation' (Preacher & Hayes, 2008).

The moderation hypotheses (i.e. $H_0$5 and $H_0$7) were tested with the help of Moderation analysis using Model 1 in PROCESS macro for SPSS (developed by Prof. Andrew F. Hayes; Hayes, 2013). To examine the statistical significance of the differential effect, Bias-corrected Bootstrapping resampling methods were used. The guidelines of Preacher et al. (2007) and Hayes (2013) were followed to perform the bootstrapping resampling procedures (on 10,000 Bootstrap sub-samples) that produced 95% bias-corrected confidence intervals. The differential (moderation) effect was considered statistically significant at α=.05 when the corresponding bias-corrected lower and upper limit of 95% confidence intervals of the effects did not include zero (Field, 2013).

Further, follow up analyses as prescribed by Aiken, West, and Reno (1991) were performed by plotting the outcome variables against the predictor variables, separately for male and female students. Further, simple slope analyses were performed to examine whether the slopes of the regression lines differed significantly for different gender. Further, the size of the interaction (moderation) effect was calculated in terms of f-squared (f^2) statistic (Selya, Rose, Dierker, Hedeker, & Mermelstein, 2012): where $f^2 \geq 0.02$ implies small effect, $f^2 \geq 0.15$ implies medium effect, and $f^2 \geq 0.35$ implies a large effect (Cohen, 1988).

3.6 Procedures for developing research tools

3.6.1 Development and standardization of student engagement in learning scale (SELS)

Student engagement acts as one of the protective factors for students' dropout from school (Klem & Connell, 2004), underachievement (D'Mello, Dieterle, & Duckworth, 2017) and their deviant behavior in class (Wang & Fredricks, 2014). Besides, student engagement plays the role of a constructive element for students' learning (Ladd & Dinella, 2009), their satisfaction in life (Ko, 2011), and their academic achievement

(Appleton, Christenson, Kim, & Reschly, 2006). Thus, student engagement becomes a promising variable to elevate the outcomes of school education.

Earlier student engagement was conceptualized to be a two-dimensional construct with the dimensions: psychological and behavioral engagement (Willms, 2003). Later cognitive dimension was added to make it a three-dimensional construct (Wang & Fredricks, 2014). Earlier, student engagement was considered consisting three sub-scales: academic, cognitive, behavioral, and psychological engagement (Reschly & Christenson, 2006). Further, some researchers also proposed a new dimension as agentic engagement (Reeve & Tseng, 2011). However, it has been suggested that extensive researches are required for robust validation of the four-dimensional structure of student engagement with the new dimension: agentic engagement (Sinatra, Heddy, & Lombardi, 2015). Nevertheless, prior studies have considered following four engagement sub-scales: academic engagement, affective engagement, social engagement with teachers, social engagement with peers, and cognitive engagement (Zhoc, Webster, King, Li, & Chung, 2018). Further, Pekrun and Linnenbrink-Garcia (2012) suggested following five dimensions of student engagement: cognitive, motivational, behavioral, social-behavioral, and cognitive-behavioral engagement. Further, eight-dimensional (Gordon et al., 2008) measurement models were also proposed. Therefore, the structural model of student engagement in terms of number of sub-scales (dimensions) still remains under debate.

Further, another issue regarding student engagement is that the items in the instruments measuring engagement are either related to antecedents or consequences of engagement. Hence, the validity of those instruments gets contaminated due to inclusion of the antecedents and consequences of the construct itself (Sinatra et al., 2015). For example, researchers have considered attendance (Appleton et al., 2006), problem-solving ability, and other academic competence (Wang, Willett, & Eccles, 2011) as the indicators of student engagement although those are mere antecedents of student engagement. In other studies, consequences of student engagement for example, academic records and behavioral features of the students (e.g. regular completion of home works), teacher-report of classroom behavior, and credits earned by the students (Lehr, Sinclair, &

Christenson, 2004) have been regarded as the indicators of engagement. Hence, those instruments fail to represent engagement as a process (Csiksentmihalyi & Larson, 1987). This results in imperfection and inaccuracy in the measurement of engagement.

Eventually, there is a need to develop an instrument with items representing engagement as a dynamic process that acts as a regulatory mechanism of enhancing the quality of the classroom teaching-learning process and its outcomes. The measure of student engagement should incorporate students' opinions of contextual impacts on engagement as well as engagement indicators in order to better comprehend the fundamentally subjective experience of students' involvement in learning.

Furthermore, in order to compare groups (e.g. boys and girls) using latent mean scores, the measure must have a 'strong' measurement invariance (i.e. identical factor loadings and equal intercepts; Byrne, 2008). Measurement invariance refers to the degree to which the measuring tool's parameters are comparable across groups (Byrne, 2008). Multi-group CFA is commonly used to investigate measurement invariance. It is assumed that the factorial structure of the same latent construct(s) stays constant across all groups (Byrne & van de Vijver, 2010). However, if the equivalence or invariance of an assessment instrument is not maintained, the inferences derived from the analysis of data obtained using the instrument may be faulty (Byrne, 2008), and the conclusions based on group comparisons will be incorrect.

3.6.1.1 Item writing

The construction of the items in the instrument was primarily based on a thorough review of published tools (Greene, Miller, Crowson, Duke, & Akey, 2004; Wolters, 2004; see Table 3.2 to 3.5) measuring the student engagement construct. The intent was to gather the range, content coverage, and varieties of items. For this, various tools on engagement construct were reviewed and an item bank with 94 items was developed.

Further, five anchor points starting from 'strongly disagree' (1) to 'strongly agree' (5) were assigned to rate the scale items. Since seven-point Likert scales are lengthy and might lead respondents to get confused, five-point Likert scales were used to lessen respondents' annoyance and thus, enhance response rates (Pai & Huang, 2011). The items

were representative of different dimensions of student engagement identified in the earlier studies that include different types of student engagement, for example, school/campus engagement, classroom engagement, engagement in online/blended learning, lesson/subject-specific engagement and uni-dimensional measure of student engagement (see Table 3.2 to 3.5).

3.6.1.2 Initial Try-out:

Items that contributed just as antecedents and consequents of engagement were removed through a planned iterative procedure. The initial try-out was conducted to ensure that no items representing either antecedents or consequences remain in the preliminary draft, items those represent student engagement as a process are retained in the draft, items present in the tools covered all the areas of the construct, and the items are consistent with the context of the present study.

A. Expert Try-out:

The preliminary draft of the target instrument containing 94 scale items was sent to fifteen experts (Lambie et al., 2017) for their suggestions and recommendations. The scholars were selected as experts whose area of specialization is measurement and evaluation in Psychology and Education and who have minimum 10 years of experience in teaching and research in those specified fields (Ikart, 2019).

In view of the criticism and comment by the experts, 32 items were removed from the Student Engagement in Learning Scale while others were modified accordingly. Hence, a final version including 62 items (28 negative items) was compiled based on the valuable comments from experts. Then, this draft was also approved by the Departmental Research Committee (DRC), Department of Education, Tezpur University. This allowed the researcher to conduct the pre-test (individual try-out) of the tool.

Table 3.2

Major dimensions of campus/school engagement, their definition/characteristics, sample items and their related sources.

Dimensions	Definition/characteristics	Sample items	Sources of the items
Valuing	Affective connections to the institution	'University is of great importance in my life'.	Gunuc & Kuzu (2015), Finn (1993), Voelkl (1996)
Sense of belonging	Affective involvement is measured by a sense of belonging. It is a student's good attitude toward his or her institution.	'I feel secure in campus'.	Gunuc & Kuzu (2015), Appleton et al. 2006; Sutherland, 2010; Singh and Srivastava, 2014
Participation	Participation is also known as social engagement and is defined as taking part, particularly in campus or out-of-class activities.	'I go to campus willingly'.	Willms (2003), Gunuc & Kuzu (2015)
Affective	Students' feelings and sense of belonging to school and coursework, as well as their views about learning, instructors, academics, and peers	'My school is a place where I make friends easily'.	Appleton et al., 2006; Gutiérrez, et al., 2016; Veiga, 2016; Whitney et al., 2019
Agency	Students' positive participation in academic learning at school.	'I make suggestions to teachers about how to improve classes'.	Gutiérrez, et al., 2016; Veiga, 2016

Table 3.3

Major dimensions of classroom engagement, their definition/characteristics, sample items and their related sources.

Dimensions	Definition/characteristics	Sample items	Sources of the items
Cognitive	Students' approaches to and knowledge of their own learning are linked to cognitive engagement. It pertains to learning investment, learning value, learning motivation, learning objectives, self-regulation, and planning.	'I motivate myself to learn'.	Appleton et al. (2006), Lehr et al. (2004), Sutherland (2010), Fredricks et al, 2004; 2005, Gunuc & Kuzu (2015); Iqbal, Qureshi, & Asghar (2021), Dogan (2014); Whitney et al., 2019
Behavioral	Students' behavioural engagement encompasses their participation in academic, out-of-class educational activities, their efforts, as well as their attendance and participation in courses.	'I am an active student in class'.	Appleton et al. (2006), Finn (1993), Krause and Coates (2008), Lehr et al. (2004), Newmann, Wehlage, and Lamborn (1992), Gunuc & Kuzu (2015); Dogan (2014); Whitney et al., 2019; Fredricks et al, 2004; 2005
Emotional	Emotional engagement refers to students' emotional reactions to the teacher/staff, classmates, course material, and class, which include their attitudes, interests, connections, and values. Furthermore,	'My teachers are always near me when I need them'.	Appleton et al. (2006), Fredricks et al, 2004; 2005, Kember, Lee, and Li (2001), Sutherland (2010), Gunuc & Kuzu (2015);

	sentiments like as belonging to class, enjoying the lesson, and being a part of a group are regarded within the purview of emotional engagement.		Dogan (2014)
Agentic	The ability to alter and modify interactive learning processes is associated with agentic engagement.	'I let my teacher know what I am interested in'	Mameli & Passini, 2019; Reeve, 2013; Reeve & Tseng, 2011
Teacher-student Relationships	Care and concerns of teachers for the students and their learning	'I enjoy talking to the teachers here.'	Betts et al., 2010
Peer Support for Learning	Support, sharing and assistance from the peers	'Students here respect what I have to say.'	Betts et al., 2010
Family Support for Learning	Emotions and attachment with the family	'My family/guardian(s) are there for me when I need them.'	Betts et al., 2010
Physical engagement.	Physical effort and labor invested in learning	'I work with intensity on assignments for this class/course.'	Burch et al., 2015

Table 3.4

Major dimensions of lesson/subject-specific engagement, their definition/characteristics, sample items and their related sources.

Dimensions	Definition/characteristics	Sample items	Sources of the items
Cognitive	Reliance, Surface and deep strategy	'I solve problems according to what the teacher	Kong et al., 2003

		teaches.’	
Behavioral	Interest, achievement, anxiety and frustration	‘I really make an effort in the mathematics lesson.’	Kong et al., 2003
Affective	Attentiveness, diligence and time spent	‘I feel excited when we start a new topic in mathematics.’	Kong et al., 2003
Engagement on Science Lessons and Tasks	Interesting and meaningful science lessons and performance tasks	‘I am inspired to learn new things in Science class.’	Baraquia (2019)
Social	Attachment to the teachers and peers	‘I try to work with others who can help me in science/math.’	Wang et al., 2016

Table 3.5

Uni-dimensional measure of student engagement: definition/characteristics, sample items and related sources.

Dimensions	Definition/characteristics	Sample items	Sources of the items
Student engagement	Engagement is a broad notion that involves many distinct actions and attitudes, with researchers and theorists labelling these activities with various titles such as participation, identification, motivation, and membership.	‘I check my schoolwork for mistakes’	Cavanagh (2015); Moreira et al., 2019

B. Individual Try-out (Pre-test and Pilot study):

After Experts Tryout, the student engagement in learning scale was left with 62 items. To reduce measurement error (Blair & Conrad, 2011), pre-testing was conducted with these 62 scale items on 30 students (Perneger, Courvoisier, Hudelson, & Gayet-Ageron, 2015) of Class-X who were not a part of the final study sample (Cooper and Schindler, 2011; Kumar, Talib & Ramayah, 2013). This pre-testing was conducted in order to determine whether any item was unclear, poorly sequenced, or confusing in the scale, and whether the instructions were given properly to the respondents (Kumar et al., 2013).

Results of the pre-testing demonstrated that the scores for 8 items (e.g. I study in class as teacher praises me, whatever I learn in class will be useful in future) showed minor deviations and skewness from the mean score. Those items were excluded from the scale. Additionally, 4 items (e.g. I wait for the moment when class finishes, I do not remain absent in class without genuine reason) were deleted due to their ambiguity and lack of clarity to the respondents, whereas 7 items were dropped (e.g. I am determined to score well in examinations, I remain aware of what is being discussed in class) owing to item social desirability bias (Memon, Ting, Ramayah, Chuah, & Cheah, 2017). In sum, on the basis of individual try-out, 19 items were removed from the preliminary draft and 12 items were slightly modified. Thus, the preliminary draft of the SELS included 43 items (20 negative items).

Further, the revised questionnaire with 43 items underwent a second round of pre-testing (Memon et al., 2017) that suggested for no further modifications. Finally, a rating scale with 43 items (20 negative) was developed which was to be rated by the students based on their perceptions. Thus, the modification in the items was found to be effective. Finally, the instrument with 43 items was piloted on another 30 students (Memon et al., 2017). The scores were reversed for the negative items before conducting any calculation. Then, a reliability analysis was performed considering all the items representing a single construct. The results depicted acceptable reliability as the Cronbach alpha (α) value was 0.89. However, the factorial structure of the construct was further tested in Group Try-out phase.

C. Group Try-out:

The student engagement in learning scale with 43 items was administered on 414 tenth graders in government secondary schools in Birbhum, West Bengal. Among them 196 (47.343%) were males and 218 (52.657%) were females. Consents were taken from the head of the institutions for getting the responses from the students. Then, each student was provided with one copy of the student engagement in learning scale and was asked to follow the instructions. Further, they were informed that their responses will neither be evaluated nor be disclosed anywhere.

The items were assigned with the scores as '5' to '1' for 'Strongly Agree' to 'Strongly Disagree', respectively. However, the reverse scoring procedure was applied for the negative items. The total score of each respondent was calculated by adding the scores in each engagement sub-scale. Lesser values of total scores indicated lower level of engagement whereas the higher values of the total scores indicated higher level of engagement.

Further, item analysis was performed for 43 scale items. The raw scores of those items those were obtained in the form of the responses from 414 students were used for item analysis. Firstly, from the responses obtained from Group Tryout, individual scores were calculated and were arranged in descending order. Then, respondents belonging to the top 27% group (N= 112) and bottom 27% (N= 112) group were identified. Then, the data of 112 subjects belonging to the top 27% group and 112 belonging to the bottom 27% group were analyzed for 43 scale items to calculate the Discrimination Index of the items (Table 3.6):

Table 3.6

Item-wise M, SD, and t-values

N= 224

Item No.	M_1	SD_1	M_2	SD_2	t-value	p-value	Remark
1	3.04	0.72	3.34	0.68	-3.24	0.00	Selected
2	3.10	0.86	3.26	0.79	-1.46	0.15	Rejected

3	3.05	0.85	3.27	0.72	-2.04	0.04	Selected
4	3.13	0.90	3.31	0.74	-1.71	0.09	Rejected
5	3.19	0.80	3.25	0.72	-0.62	0.54	Rejected
6	3.09	0.83	3.34	0.82	-2.26	0.02	Selected
7	3.29	0.88	3.59	0.84	-2.63	0.01	Selected
8	3.15	0.84	2.96	0.68	1.83	0.07	Rejected
9	3.22	0.91	2.97	0.68	2.34	0.02	Selected
10	3.12	0.82	2.99	0.66	1.25	0.21	Rejected
11	3.21	0.81	3.03	0.72	1.75	0.08	Rejected
12	3.10	0.80	3.04	0.72	0.61	0.54	Rejected
13	3.30	0.90	3.01	0.68	2.77	0.01	Selected
14	3.31	0.92	2.97	0.66	3.16	0.00	Selected
15	3.20	0.92	3.02	1.05	1.36	0.18	Rejected
16	3.14	0.76	3.00	0.70	1.47	0.14	Rejected
17	3.24	0.82	2.90	0.71	3.31	0.00	Selected
18	3.21	0.74	2.81	0.90	3.66	0.00	Selected
19	3.36	0.75	3.22	0.95	1.18	0.24	Rejected
20	3.33	0.76	3.08	1.05	2.04	0.04	Selected
21	3.46	0.84	3.28	1.02	1.51	0.13	Rejected
22	3.38	0.89	3.07	1.04	2.35	0.02	Selected
23	3.18	0.93	3.23	0.66	-0.50	0.62	Rejected
24	3.19	0.80	2.68	0.76	4.87	0.00	Selected
25	3.09	0.83	2.96	0.89	1.16	0.25	Rejected
26	2.72	1.27	3.35	1.14	-3.87	0.00	Selected
27	2.76	0.85	2.96	0.68	-1.99	0.05	Selected
28	2.97	0.94	2.97	0.68	0.00	1.00	Rejected
29	2.98	0.86	2.89	0.68	0.86	0.39	Rejected
30	3.21	0.81	2.77	0.79	4.09	0.00	Selected
31	2.79	0.80	2.99	0.70	-1.96	0.05	Rejected
32	3.24	0.82	2.18	1.03	8.53	0.00	Selected
33	3.21	0.74	2.49	0.78	7.10	0.00	Selected
34	3.36	0.75	3.17	0.87	1.73	0.08	Rejected
35	2.63	1.27	3.08	1.05	-2.92	0.00	Selected
36	2.88	0.94	3.28	1.02	-3.01	0.00	Selected
37	2.95	0.98	3.04	1.04	-0.72	0.47	Rejected
38	3.04	0.72	2.58	1.21	3.41	0.00	Selected
39	3.05	0.85	2.75	1.14	2.26	0.02	Selected
40	3.13	0.90	3.18	0.81	-0.47	0.64	Rejected
41	2.88	1.09	2.97	0.66	-0.81	0.42	Rejected
42	3.13	1.01	2.63	1.25	3.29	0.00	Selected
43	3.14	0.76	2.65	0.72	4.97	0.00	Selected

From the Table 3.6, it is clear that t-values for the items at serial numbers 1, 3, 6,7, 9, 13, 14, 17, 18, 20, 22, 24, 26, 27, 30, 32, 33, 35, 36, 38, 39, 42, 43 are significant either at 0.01 level or at 0.05 level. These items were found to discriminate between the respondents belonging to the top 27% group and the bottom 27% group. Thus, these 23 scale items (11 negative items) were kept in the instrument for the Final Try-out.

3.6.1.3 Final Try-out:

Data were collected from two independent randomized sub-samples for conducting EFA and CFA. The first sub-sample contained 576 (292 females) and second sub-sample contained 631 (327 females) secondary school students in the seven districts (Cooch Behar, Purba Bardhaman, Nadia, Murshidabad, Birbhum, North Dinajpur and Hooghly) of West Bengal. None of the samples was included in the final data collection of the study. Bengali was the medium of instruction and communication in all the selected schools. Consents were taken from both the head of the institutions and of the students for getting the responses. Then, each student was provided with one copy of the SELS and was asked to follow the instructions.

A. Preliminary descriptive analyses

Before doing the main analyses, data were examined to check whether the statistical assumptions were fulfilled. There was no missing value in the dataset. Measures of central tendency and variability were performed for each scale item (see Table 3.7). Further, to check whether the data followed the normal probability distribution, skewness and kurtosis were checked for all items. Results (see Table 3.7) showed that the values for all items were within the statistically acceptable range (skewness<2 and kurtosis<7; Curran, West, & Finch, 1996).

Descriptive analysis of the scores for 23 scale items rated on the 5-point Likert scale was executed. The means and standard deviations ranged from 2.37 to 4.63 and from 0.760 to 1.780, respectively. The standardized value (Z) of the skewness and kurtosis (see Table 3.7) for all items were within the statistical thresholds (Kline, 2011). Further, the item-total correlations and the inter-item correlations for each scale item were calculated and the results have been demonstrated in Table 3.8 and in Table 3.9, respectively.

Table 3.7

Mean, SD, skewness and kurtosis of the 23 scale items

Item no.	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
1	3.29	1.054	-0.254	0.091	-0.373	0.055
2	3.27	1.110	-0.276	0.080	-0.537	0.132
3	3.30	1.079	-0.284	0.081	-0.446	0.185
4	3.29	1.117	-0.324	0.093	-0.502	0.183
5	3.28	1.083	-0.313	0.096	-0.366	0.094
6	3.37	0.760	-0.668	0.098	-0.251	0.046
7	3.41	1.089	-0.707	0.097	-0.228	0.141
8	3.38	1.089	-0.384	0.087	-0.361	0.153
9	3.26	1.091	-0.036	0.084	-0.577	0.197
10	3.24	0.980	-0.079	0.079	-0.646	0.091
11	3.23	1.084	-0.203	0.067	-0.476	0.136
12	3.25	1.077	-0.193	0.050	-0.425	0.181
13	4.63	1.077	-0.151	0.071	-0.487	0.164
14	3.27	1.059	-0.397	0.079	-0.423	0.126
15	4.06	1.062	-0.316	0.075	-0.529	0.112
16	3.96	1.780	-0.238	0.098	-0.535	0.063
17	3.26	1.076	-0.319	0.096	-0.421	0.077
18	3.29	1.069	-0.335	0.091	-0.453	0.107
19	3.24	1.108	-0.32	0.094	-0.52	0.057
20	3.26	1.116	-0.405	0.088	-0.51	0.163
21	3.36	1.136	-0.63	0.074	-0.376	0.102
22	3.44	1.156	-0.623	0.068	-0.482	0.021
23	2.37	1.140	-0.513	0.081	-0.53	0.135

Table 3.8

Item-Total Statistics

Items	Corrected Item-Total Correlation	Items	Corrected Item-Total Correlation
1	0.42	13	0.48
2	0.41	14	0.48
3	0.42	15	0.44
4	0.47	16	0.48
5	0.42	17	0.52
6	0.39	18	0.53
7	0.42	19	0.56
8	0.45	20	0.52
9	0.45	21	0.48
10	0.46	22	0.49
11	0.46	23	0.47
12	0.48		

Table 3.9

Inter-item Correlations for SELS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	1																						
2	0.30	1																					
3	0.26	0.34	1																				
4	0.28	0.30	0.27	1																			
5	0.31	0.25	0.16	0.35	1																		
6	0.18	0.23	0.31	0.25	0.21	1																	
7	0.20	0.31	0.26	0.35	0.24	0.30	1																
8	0.09	0.04	0.07	0.05	0.07	0.04	0.09	1															
9	0.10	0.10	0.12	0.11	0.06	0.03	0.08	0.11	1														
10	0.09	0.08	0.09	0.10	0.04	0.01	0.06	0.16	0.28	1													
11	0.09	0.10	0.09	0.13	0.05	0.05	0.08	0.24	0.25	0.30	1												
12	0.10	0.07	0.10	0.11	0.05	0.02	0.06	0.07	0.19	0.22	0.33	1											
13	0.08	0.09	0.06	0.11	0.05	0.07	0.09	0.15	0.26	0.09	0.25	0.32	1										
14	0.08	0.10	0.06	0.10	0.06	0.05	0.10	0.21	0.26	0.06	0.28	0.21	0.04	1									
15	0.08	0.09	0.07	0.10	0.05	0.01	0.04	0.30	0.33	0.27	0.38	0.19	0.18	0.12	1								
16	0.08	0.10	0.09	0.13	0.12	0.07	0.09	0.11	0.07	0.07	0.08	0.11	0.15	0.14	0.09	1							
17	0.08	0.08	0.14	0.17	0.14	0.16	0.10	0.18	0.12	0.17	0.12	0.12	0.15	0.15	0.11	0.23	1						
18	0.12	0.14	0.16	0.19	0.19	0.19	0.17	0.16	0.08	0.13	0.12	0.13	0.18	0.16	0.14	0.15	0.26	1					
19	0.16	0.12	0.15	0.18	0.16	0.11	0.11	0.17	0.12	0.16	0.13	0.18	0.14	0.16	0.16	0.21	0.29	0.27	1				
20	0.09	0.10	0.13	0.14	0.13	0.16	0.13	0.12	0.09	0.10	0.10	0.12	0.11	0.14	0.12	0.20	0.20	0.21	0.34	1			
21	0.09	0.09	0.09	0.14	0.11	0.13	0.11	0.08	0.07	0.09	0.07	0.10	0.09	0.11	0.07	0.25	0.07	0.33	0.20	0.14	1		
22	0.11	0.07	0.11	0.16	0.13	0.14	0.14	0.10	0.09	0.10	0.05	0.08	0.09	0.07	0.05	0.27	0.27	0.36	0.21	0.13	0.25	1	
23	0.08	0.05	0.10	0.10	0.08	0.13	0.15	0.12	0.10	0.08	0.11	0.14	0.10	0.14	0.09	0.16	0.21	0.19	0.27	0.07	0.27	0.22	1

B. Main analyses

Data analysis was performed using SPSS 26.0 and AMOS 23.0 module. Firstly, to develop the measurement model for student engagement EFA was performed with the 23 scale items. Reverse scoring was carefully done for the negative items before performing any analysis.

Next, to test and evaluate the measurement model developed using EFA, a series of Confirmatory Factor Analyses (CFA) was performed. Firstly, a zero-order 1-factor model was evaluated by taking the entire set of 23 items together and loaded on a single factor. Next, three 2-factor models were tested taking anyone out of the three factors (viz. cognitive, behavioral, and emotional engagement) as a distinct factor and combining the items of the rest of the two factors.

Finally, a first-order 3-factor model was evaluated. The 5 measurement models were then compared on the basis following model fit indices: $\chi 2$ statistic and associated p-value, CFI, GFI, PCFI, PNFI, RMSEA, and SRMR, where the threshold values for CFI and GFI was ≥.90 (Bentler & Bonnet, 1980), ≥.50 for PCFI and PNFI (James, Mulaik, & Brett, 1982), and <.80 for acceptable (MacCallum, Browne, & Sugawara, 1996) or ≤.60 (Hu & Bentler, 1999; Vanderberg & Lance, 2000) for RMSEA and SRMR for a good fit. Further, the construct validity was assessed from the results of CFA.

Besides, the mean scores of student engagement dimensions were compared to evaluate whether the evaluation of the factors differs across gender of the students. The factorial invariance of the instrument was evaluated using multi-group confirmatory factor analysis (MGCFA) (Byrne, 2010). Finally, inter-correlations among the dimensions of the construct were also calculated using Pearson's product-moment correlations.

I. Development of the measurement model: Exploratory Factor Analysis

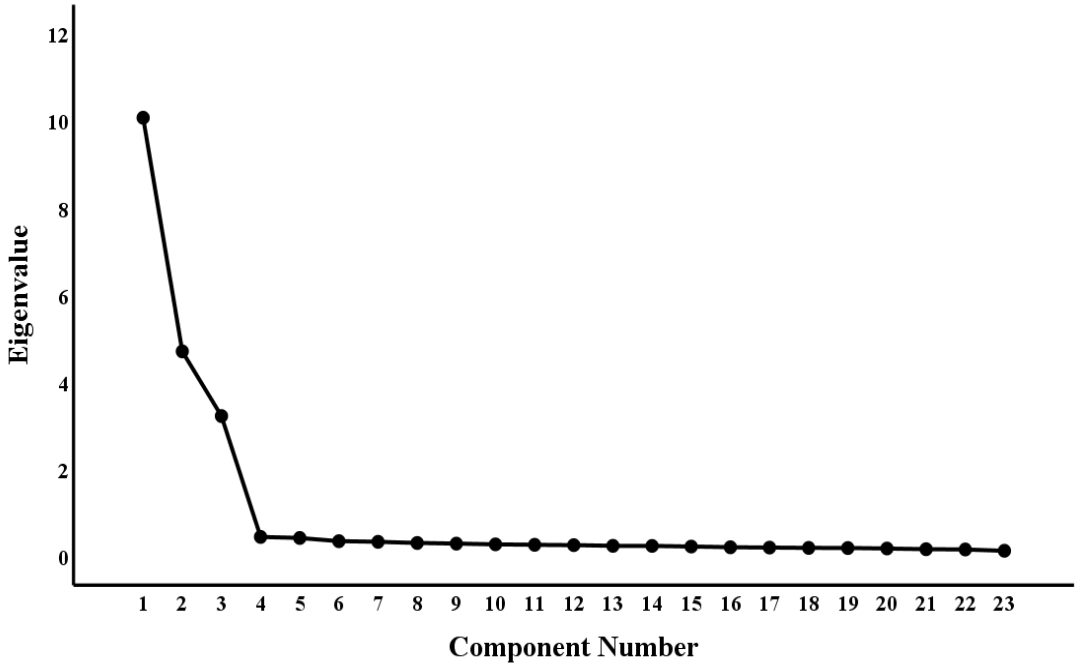
To explore the factors of student engagement, EFA was performed using promax rotation and principal components analysis as the method of extraction of factors. The oblique rotation used to allow that the factors of the construct are related (Byrne, 2008). Before that, the prerequisites for conducting EFA were checked. Firstly, Kaiser-Meyer-Olkin's

measure of sample adequacy (.960) was greater than 0.6 and Bartlett's Test of Sphericity was significant ($\chi 2$=13450.475, df=253, p<.001). Further, the non-zero determinant value ($4.881 \times e^{-11}$) of the correlation matrix satisfied the criteria of positive definiteness.

Results of the EFA came up with three-factor solution following Kaiser Criterion (eigenvalue>1; Cattell, 1966). The examination of the scree plot was supported the three-factor solution (see Graph 3.1). Further, to determine the number of factors to be retained in the measurement model, Parallel analysis (O'Connor, 2000) was run. The comparison between the eigenvalues of raw data from EFA with 95th percentile of random eigenvalues suggested retaining three factors (see Table 3.10). This can be visualized in Graph 3.2. Additionally, Wayne Velicer's Minimum Average Partial (MAP) test (Velicer, 1976; Velicer et al., 2000) indicated that the three-factor structure of student engagement construct achieved both a minimum of 0.01 average partial and the lowest Bayesian information criterion (BIC) as suggested from the parallel analysis.

Graph 3.1

Scree plot for the 3-factor solution of EFA



Graph 3.2

Tsplot of Raw Data Eigenvalues, & Mean & (95th) Percentile Random Data Eigenvalues

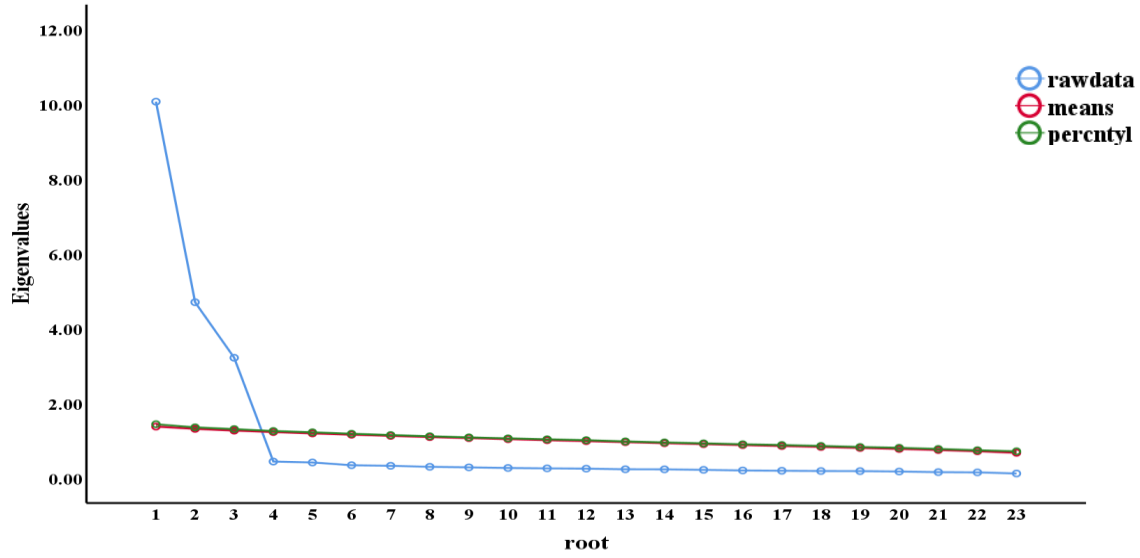


Table 3.10

Raw Data Eigenvalues, & Mean & (95th) Percentile Random Data Eigenvalues

Root	Raw Data Eigenvalues	Means of Eigenvalues	(95th) Percentile Eigenvalues
1	10.072	1.378	1.441
2	4.705	1.316	1.357
3	3.219	1.270	1.310
4	0.443	1.231	1.260
5	0.418	1.194	1.225
6	0.345	1.161	1.187
7	0.330	1.129	1.152
8	0.303	1.097	1.120
9	0.288	1.069	1.090
10	0.272	1.042	1.064
11	0.261	1.015	1.039
12	0.254	0.988	1.015
13	0.237	0.962	0.982
14	0.236	0.934	0.955
15	0.221	0.909	0.930
16	0.206	0.883	0.906
17	0.198	0.858	0.885
18	0.190	0.834	0.860
19	0.188	0.806	0.833
20	0.177	0.779	0.811
21	0.160	0.751	0.781
22	0.154	0.718	0.746
23	0.123	0.676	0.716

The measurement model explained 78.19% of the variance in the latent construct in total. The rotated component matrix showed that the standardized factor loadings for all the items on corresponding latent factors were greater than |.45| (Hair, Tatham, Anderson, and Black, 1998) (see Table 3.11). The item-total statistics were also calculated (see Table 3.8).

The items were assigned to each scale dimension following their loading patterns. The nomenclature of the dimensions was based on the overall nature of the included items for each corresponding dimensions. The three-factor model was developed: Cognitive engagement (CE) (item no.: 1 to 7 [CE1 to CE7]), Behavioral engagement (BE) (item no.: 8 to 15 [BE1 to BE8]), Emotional engagement (EE) (item no.: 16 to 23 [EE1 to EE8]) (see Table 3.11).

The first subscale (cognitive engagement) accounted for 26.5% of the total variance in student engagement. Sample items include: “Performing well in school is important to me” and “I plan to continue for higher studies”. The second subscale i.e. behavioral dimension accounted for 26.1% of the total variance in the construct. This factor included items such as “I ask the teacher for clarification” and “I remain attentive in the class”. Finally, the third subscale (emotional engagement) accounted for 25.5% of the variance. Some of the sample items are: “I feel regretted when I miss any class” and “The classes are boring”.

Finally, Cronbach’s alpha values were found to be 0.960 for cognitive, 0.963 for behavioral, 0.957 for emotional engagement, and 0.944 for the overall scale that indicates high reliability of the measurement scale along with its dimensions.

To examine whether the dimensions of student engagement are significantly related to each other, Correlational analyses (Pearson’s product-moment correlation) were performed. However, all bivariate correlation coefficients were positive and statistically significant.

Table 3.11

Rotated Component Matrix: Standardized factor loadings of 23 items in three dimensions of the student engagement construct

Scale items	Components 1	2	3
Cognitive engagement items			
1. When I am in class, I forget everything else around.	.860		
2. Performing well in class is not important to me.*	.863		
3. I plan to continue for higher studies.	.879		
4. I try to relate classroom discussions to the things I already know.	.894		
5. I try to generate examples on the concepts discussed in the class.	.845		
6. I keep on studying regularly rather than waiting for examinations.	.865		
7. Whenever I find something difficult, I give up right away.*	.859		
Behavioral engagement items			
8. I wait for the moment when class finishes.*		.856	
9. I feel sleepy in class.*		.867	
10. I ask the teacher for clarification.		.875	
11. I remain attentive in class.		.904	
12. I often whisper in class.*		.847	
13. I do not ask any questions in class.*		.845	
14. I discuss on studies with my classmates.		.871	
15. I pretend my teacher that I am working properly in the class.*		.836	
Emotional engagement items			
16. I eagerly wait for teacher's arrival in the class.			.835
17. I do not like my teacher.*			.812
18. It is better to read at home than attending class.*			.866
19. I feel regretted when I miss any class.			.853
20. The classes are boring.*			.864
21. Learning in the class is full of joy.			.836
22. My friends help me when I miss any class.			.812
23. I feel happy when teacher remains absent.*			.847
Factor inter-correlations			
Factor 1	-	.239**	.333**
Factor 2		-	.456**
Factor 3			-

Note. Correlation coefficients ≤0.45 are not shown *Negative scale items, **p<.001

The results showed that cognitive engagement showed slightly higher correlation with emotional dimension (r =0.333, p <0.001, effect size= 0.11) than with behavioral dimension (r =0.239, p <0.001, effect size= 0.06) (see Table 3.11). Further, the correlation between behavioral engagement and emotional engagement was strongest (r =0.456, p <0.01, effect size= 0.21) (see Table 3.11). This shows that the dimensions are interrelated among each other. However, the low values of effect size (Cohen, 1988) signify that those intra-construct correlations were practically small in size. This implies that even the dimensions are significantly related to each other, those measure different aspects of the student engagement construct. Hence, those are considered to be three distinct constructs.

The final draft of Student Engagement in Learning Scale consisted with 23 items (11 positively worded and 12 negatively worded). The dimension-wise distribution of serial number-wise items has been depicted in Table 3.12.

Table 3.12

Serial Number-wise distribution of items in each dimension and types of items

Sl. No.	Dimensions	Nature of items	Sl. No. of items in Final Scale	Total no. of items	Total
1	Cognitive Engagement	Positive	1, 3, 4, 5, 6	5	7
		Negative	2, 7	2	
2	Behavioral engagement	Positive	10, 11, 14	3	8
		Negative	8, 9, 12, 13, 15	5	
3	Emotional Engagement	Positive	16, 19, 21, 22	4	8
		Negative	17, 18, 20, 23	4	
Positive items= 12 + Negative items= 11				Total items	23

II. Confirmation of the measurement model: Confirmatory Factor Analysis

Results of the zero-order CFA considering all 23 items as a single-factor (1-factor model) showed poor model-fit indices: $\chi 2$= 8797.26, p<0.001, $\chi 2$/df = 38.25, CFI = 0.432, GFI= .265, PCFI= .394, PNFI= .389, RMSEA = 0.235, and SRMR = 0.272. Similarly, none of

the competing 2-factor models were supported for poor model fit resulted from a series of first-order CFA (see Table 3.13). Finally, a first-order CFA (Byrne, 2005) with the final 23 scale items under the three factors (CE, BE and EE) of student engagement was conducted. The model fit indices satisfactorily supported the 3-factor model: $\chi 2$= 405.26, p<0.001, $\chi 2$/df = 1.79, CFI = 0.99, GFI= .945, PCFI= .893, PNFI= .878, RMSEA= 0.032, SRMR = 0.029. Therefore, the three-factor model (see Figure 3.5) outperformed both the single-factor as well as the two-factor models. Finally, the 3-factor model of student engagement was retained as it showed better fit compared to other competing 1-factor model and 2-factor models. The standardized factor loadings of each item to their corresponding factors are given in Table 3.14.

Table 3.13

Model fit indices of 1-factor, 2-factor, and 3-factor models from confirmatory factor analyses

	χ^2	p	χ^2/df	CFI	GFI	PCFI	PNFI	RMSEA	SRMR
One factor model									
Model 1: CE/BE/EE[a]	8797.26	<.001	38.25	.432	.265	.394	.389	.235	.272
Two factor models									
Model 2: CE+BE/EE	4148.37	<.001	18.12	.741	.445	.674	.663	.159	.205
Model 3: BE+CE/EE	4751.98	<.001	20.75	.701	.425	.637	.627	.171	.220
Model 4: EE+CE/BE	5048.89	<.001	22.05	.681	.422	.619	.610	.176	.260
Three factor model									
Model 5: CE+BE+EE	405.26	<.001	1.79	.990	.945	.893	.878	.032	.029

[a]*Note.* A single engagement scale (either CE or BE or EE) refers to a distinct factor with the items solely loaded on that factor. A set of engagement scales that appears in a group (e.g. CE/BE/EE) refers to a common combined factor with all items from those factors loaded on that single composite factor. The + sign was used to mean distinct factors and the / sign was used to mean the combination of items in two or more factors

Figure 3.5

Three-factor measurement model of student engagement construct

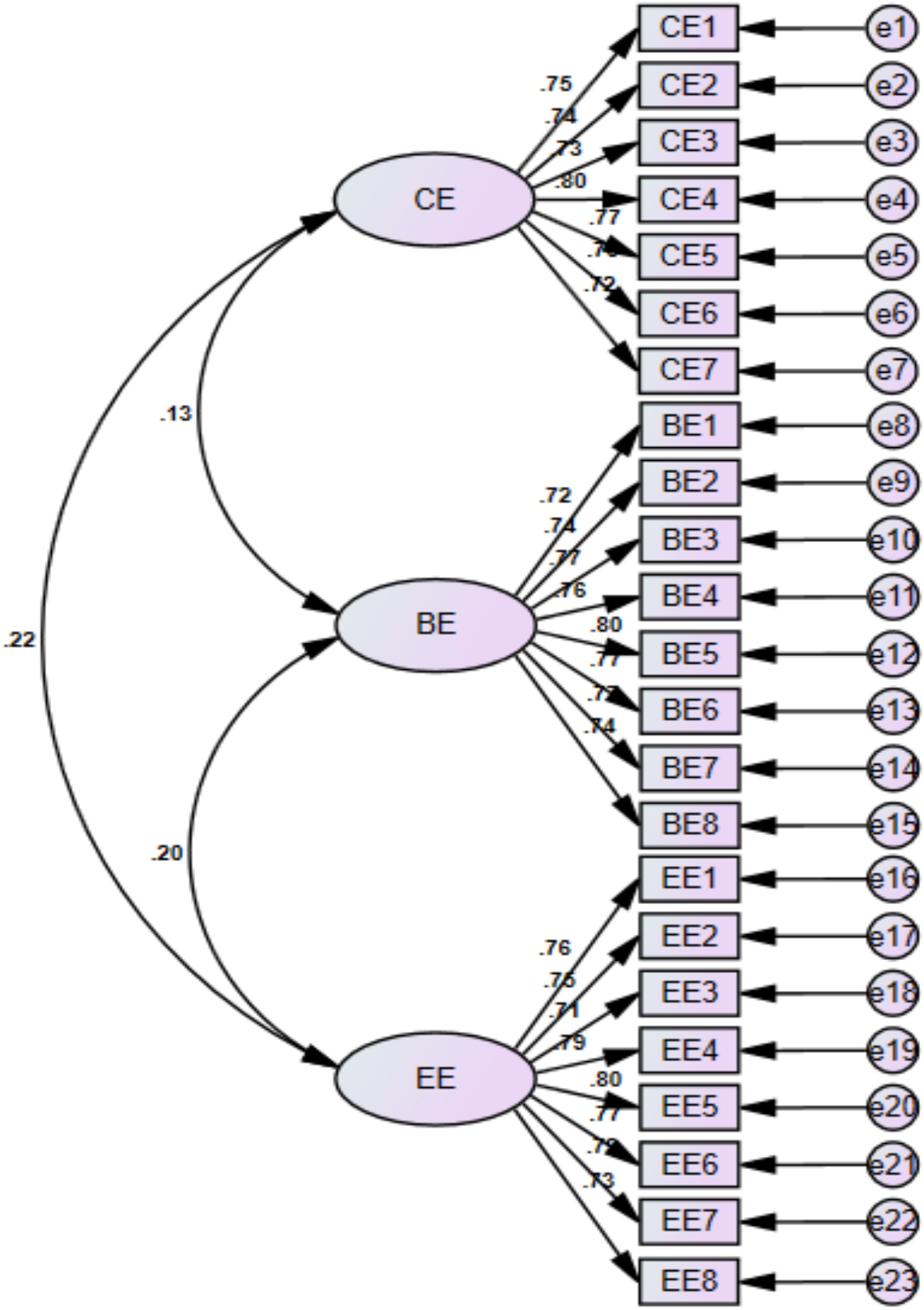


Table 3.14

Standardized Regression Weights

Regression path	Estimate	Regression path	Estimate	Regression path	Estimate
CE1<---CE	.749	BE1<---BE	.715	EE1<---EE	.756
CE2<---CE	.744	BE2<---BE	.744	EE2<---EE	.754
CE3<---CE	.733	BE3<---BE	.769	EE3<---EE	.714
CE4<---CE	.795	BE4<---BE	.759	EE4<---EE	.786
CE5<---CE	.774	BE5<---BE	.802	EE5<---EE	.803
CE6<---CE	.701	BE6<---BE	.770	EE6<---EE	.774
CE7<---CE	.719	BE7<---BE	.770	EE7<---EE	.789
		BE8<---BE	.745	EE8<---EE	.726

III. Evaluation of the psychometric properties of SELS

a. Validity

The scale has Face Validity as the scale was examined by the experts and the scale was modified accordingly based on their suggestions. For content validity, item analysis was conducted and only the items demonstrating statistically significant discriminative power were retained. Further, the Factorial Validity of the scale was also examined using First-order Confirmatory Factor Analysis. The results are given in Table 3.15 and Table 3.16.

i. Convergent validity:

The relationship between the observed variables and their corresponding latent variables was examined using convergent validity where the standardized factor loadings must be ≥0.55 (Hair, Hult, Ringle, & Sarstedt, 2017). In the three-factor model all the standardized factor loadings were statistically significant ($p<0.001$) and ranged between 0.820 and 0.931. Further, the composite reliability (CR) coefficients for each dimension of student engagement (cognitive, behavioral, and emotional engagement was 0.897, 0.916, and 0.918, respectively; see Table 3.15) demonstrated a satisfactory value above 0.7 (Fornell, 1982). Whereas, the average variance extracted (AVE) for each dimension (cognitive, behavioral, and emotional engagement was 0.556, 0.577, and 0.583, respectively; see Table 3.15) was greater than 0.50 (Fornell, 1982). These results indicate that each of the dimensions qualifies for convergent validity (CR>.7, AVE>.5, CR>AVE; Hair et al., 2017).

Table 3.15

Convergent validity of the engagement sub-scales

Sub-scales	CR	AVE	CR>0.7	AVE>0.5	CR>AVE	Convergent Validity
CE	0.897	0.556	Satisfied	Satisfied	Satisfied	Established
BE	0.916	0.577	Satisfied	Satisfied	Satisfied	Established
EE	0.918	0.583	Satisfied	Satisfied	Satisfied	Established

ii. Discriminant validity:

Further, maximum shared variance (MSV) (for CE=0.048, BE=0.040 and EE=0.048; see Table 3.16) and average shared variance (ASV) (for CE=0.031, BE=0.027 and EE=0.044; see Table 3.16) for each latent construct were calculated from intra-construct correlation coefficients in the 3-factor measurement model. Both MSV and ASV were found to be numerically less than AVE for each latent factor. Hence, all the factors qualify for discriminant validity (Hair, Black, Babin, Anderson, & Tatham, 2014).

Table 3.16

Discriminant validity of the engagement sub-scales

Sub-scales	AVE	MSV	ASV	AVE>MSV	AVE>ASV	Discriminant Validity
CE	0.556	0.048	0.031	Satisfied	Satisfied	Established
BE	0.577	0.040	0.027	Satisfied	Satisfied	Established
EE	0.583	0.048	0.044	Satisfied	Satisfied	Established

b. Reliability

The internal consistency of the Student Engagement in Learning Scale was calculated using Cronbach's Alpha (Cronbach. 1951) for each sub-scale and for the overall scale based on the scores of 631 respondents. The reliability coefficients are given in Table 3.17. The Cronbach's α value for cognitive, behavioral, and emotional engagement was 0.901, 0.920, and 0.923, respectively, and 0.884 for the overall scale was higher than the statistically acceptable figures (α>.7; Hair et al., 2017; see Table 3.17). This showed the high reliability of the instrument. Besides, the Split-Half Reliability of the scale was

obtained employing odd-even method on the scores of 631 students. The scores of the students on odd numbered items and on even numbered items were put into the Spearman-Brown Prophecy formula to obtain the Split-Half Reliability Coefficient of 0.89 for the Student Engagement in Learning Scale, which is significant at 0.01 level.

Table 3.17

Reliability coefficient of the overall scale and dimensions of Student Engagement in Learning Scale

Student Engagement Dimensions	Number of items	Cronbach's α
1. Cognitive Engagement	7	0.901
2. Behavioral Engagement	8	0.920
3. Emotional Engagement	8	0.923
Overall Student Engagement in Learning Scale	23	0.884

Therefore, the results of CFA confirmed the construct validity and internal consistency reliability of the three-factor measurement model of student engagement as obtained from the results of the EFA. The present study upholds the factorial structure of the engagement construct as proposed by the earlier researchers and confirms the absence of the fourth factor as advocated by other researchers.

c. Norms

In Final Tryout, the reliability and validity of the scale was calculated and norms of the scale were developed. The final form of the Student Engagement in Learning Scale with three key dimensions and 23 scale items were administered on a randomly selected 631 secondary school students studying in 10^{th} grade in the government schools under West Bengal Board of Secondary Education (WBBSE). The schools were situated in the seven districts (Cooch Behar, Purba Bardhaman, Nadia, Murshidabad, Birbhum, North Dinajpur and Hooghly) of West Bengal. After scoring (see Table 3.18) the 631 answer sheet, the statistical results obtained are as given in Table 3.19:

Table 3.18

Scoring system

Nature of item	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Positive	5	4	3	2	1
Negative	1	2	3	4	5

Table 3.19

Statistical results

Sl. No.	Sub-scales	N	M	SD
1	Cognitive engagement	631	22.50	5.18
2	Behavioral engagement	631	25.03	5.83
3	Emotional engagement	631	25.59	6.07
	Overall scale	631	73.12	11.41

On the basis of the statistical results presented in Table 3.19, z-Score Norms dimension-wise and for the overall scale have been developed and presented as: Dimension-wise Norms in Table 3.20, Table 3.21 and Table 3.22 whereas for the overall scale in Table 3.23, respectively. Norms for interpretation the level of each dimension and overall Student Engagement in Learning Scale have been presented in Table 3.24.

Table 3.20

z-score norms for cognitive engagement sub-scale

M= 22.50		SD= 5.18			N= 631
Raw score	Z-score	Raw score	Z-score	Raw score	Z-score
8	-2.80	17	-1.06	26	+0.68
9	-2.61	18	-0.87	27	+0.87
10	-2.41	19	-0.68	28	+1.06
11	-2.22	20	-0.48	29	+1.25
12	-2.03	21	-0.29	30	+1.45
13	-1.83	22	-0.10	31	+1.64
14	-1.64	23	+0.10	32	+1.83
15	-1.45	24	+0.29	33	+2.03
16	-1.25	25	+0.48	34	+2.22

Table 3.21

z-score norms for behavioral engagement sub-scale

M= 25.03 SD= 5.83 N= 631

Raw score	Z-score	Raw score	Z-score	Raw score	Z-score
8	-2.92	18	-1.21	28	+0.51
9	-2.75	19	-1.03	29	+0.68
10	-2.58	20	-0.86	30	+0.85
11	-2.41	21	-0.69	31	+1.02
12	-2.23	22	-0.52	32	+1.20
13	-2.06	23	-0.35	33	+1.37
14	-1.89	24	-0.18	34	+1.54
15	-1.72	25	-0.01	35	+1.71
16	-1.55	26	+0.17	36	+1.88
17	-1.38	27	+0.34	37	+2.05

Table 3.22

z-score norms for emotional engagement sub-scale

M= 25.59 SD= 6.07 N= 631

Raw score	Z-score	Raw score	Z-score	Raw score	Z-score
10	-2.57	20	-0.92	30	+0.73
11	-2.40	21	-0.76	31	+0.89
12	-2.24	22	-0.59	32	+1.06
13	-2.07	23	-0.43	33	+1.22
14	-1.91	24	-0.26	34	+1.39
15	-1.74	25	-0.10	35	+1.55
16	-1.58	26	+0.07	36	+1.71
17	-1.42	27	+0.23	37	+1.88
18	-1.25	28	+0.40	38	+2.04
19	-1.09	29	+0.56	39	+2.21

Table 3.23

z-score norms for overall student engagement in learning scale

M= 73.12		SD= 11.41			N= 631
Raw score	Z-score	Raw score	Z-score	Raw score	Z-score
34	-3.43	57	-1.41	80	+0.60
35	-3.34	58	-1.33	81	+0.69
36	-3.25	59	-1.24	82	+0.78
37	-3.17	60	-1.15	83	+0.87
38	-3.08	61	-1.06	84	+0.95
39	-2.99	62	-0.97	85	+1.04
40	-2.90	63	-0.89	86	+1.13
41	-2.82	64	-0.80	87	+1.22
42	-2.73	65	-0.71	88	+1.30
43	-2.64	66	-0.62	89	+1.39
44	-2.55	67	-0.54	90	+1.48
45	-2.46	68	-0.45	91	+1.57
46	-2.38	69	-0.36	92	+1.65
47	-2.29	70	-0.27	93	+1.74
48	-2.20	71	-0.19	94	+1.83
49	-2.11	72	-0.10	95	+1.92
50	-2.03	73	-0.01	96	+2.01
51	-1.94	74	+0.08	97	+2.09
52	-1.85	75	+0.16	98	+2.18
53	-1.76	76	+0.25	99	+2.27
54	-1.68	77	+0.34	100	+2.36
55	-1.59	78	+0.43	101	+2.44
56	-1.50	79	+0.52	102	+2.53

Table 3.24

Norms for interpretation of the levels of student engagement in learning: Dimension-wise and overall scale

S. No.	Raw score range				z-score range	Levels of student engagement
	Dimension-wise raw score range			Overall scale		
	CE	BE	EE			
1	33 & above	37 & above	38 & above	96 & above	+2.01 & above	Very high
2	30 to 32	33 to 36	34 to 37	88 to 95	+1.26 to +2.00	High
3	26 to 29	28 to 32	29 to 33	79 to 87	+0.51 to +1.25	Moderately high
4	20 to 25	23 to 27	23 to 28	67 to 78	-0.50 to +0.50	Average
5	16 to 19	18 to 22	18 to 22	59 to 66	-1.25 to -0.50	Moderately low
6	13 to 15	14 to 17	14 to 17	50 to 58	-2.00 to -1.26	Low
7	12 & below	13 & below	13 & below	49 & below	-2.01 & below	Very low

d. Factorial invariance of the student engagement measure across gender

Firstly, the unconstrained model (M_0) was evaluated to test the configural invariance of the 3-factor measurement model across the groups based on the demographic variables (i.e. gender, the locale of school, and school grade). Results (see Table 3.25) showed acceptable model fit indices indicating that the factorial structure stands invariant across all the compared groups. This model served as a reference for comparing the subsequent nested models with increasing restrictions. Secondly, the first constrained model i.e. Model 1 (M_1) (constrained by standardized factor loadings) was examined to check the equivalence in the matrix of standardized factor loadings of the different sub-samples (metric invariance).

Results (see Table 3.25) showed satisfactory fit indices for M1 where no significant increment in RMSEA and CFI (ΔRMSEA<.015, ΔCFI<.01; Cheung & Rensvold, 2002) was recorded as compared to the baseline model (M_0). This indicated that there were no significant differences between the baseline model (M_0) and the constraint model (M_1). Therefore, there were no significant differences in the pattern of standardized factor loadings indicating that the factorial structure of the construct was equivalent across the groups. The results suggested that M_0 is equivalent to M_1. Hence, the metric invariance of the measurement model was established.

Thirdly, the second nested model (M_2) constrained by intercepts was evaluated. Results (see Table 3.25) showed acceptable model fits for Model 2 (M_2) and proposed equivalence of intercepts (scalar invariance) of the different groups as the model fit indices of M_2 were found similar to the previous constrained model (M_1). While comparing the fit indices, it was found that the difference between RMSEA and CFI values did not exceed the threshold limits. Therefore, the constraint model (M_2) was equivalent for all groups and the invariance of intercepts was supported.

Finally, the third constrained model i.e. Model 4 (M_4) was proposed where the error variances and covariances were restricted to be equal among the groups (residual invariance). When Model 4 (M_4) was compared with the previous constraint model (M_3), it was found that though the constraint model (M_4) showed acceptable fit indices, the difference in RMSEA and CFI exceeded the criterion values (ΔRMSEA= .015, ΔCFI= .01). Therefore, the constraint model did not fulfill the criteria of residual invariance and, hence, error variances and covariances could not be considered equivalent across groups.

However, the measurement model of student engagement satisfied ‘strong invariance’ (Byrne, 2008) across groups. Similarly, the hierarchical steps were followed for the other groups based on the locale of school and school grade. Results (see Table 3.25) show that the measurement model of student engagement followed strong invariance across gender.

Table 3.25

Evaluation of factorial invariance of the measurement model of student engagement across students' gender

Models	χ^2	Df	χ^2/df	$\Delta\chi^2$	Δdf	CFI	RMSEA [90% CIs]	ΔCFI	ΔRMSEA
First-order 3-factor Student engagement measurement models									
M_0	719.19	454	1.584	---	---	.985	.027 [.022-.031]	---	---
M_1	750.64	477	1.574	31.45	24	.984	.027 [.023-.031]	.001	0
M_2	783.85	500	1.568	33.21	24	.982	.028 [.024-.032]	.002	0.001
M_3	790.81	503	1.572	6.96	3	.982	.028 [.024-.032]	0	0

Note. M0: fully unconstrained (Baseline) model [Configural invariance], M1: measurement model with fixed standardized factor loadings [Metric invariance], M2: measurement model with fixed standardized factor loadings and item intercepts [Scalar invariance], M3: measurement model with fixed standardized factor loadings, item intercepts and factor covariances [Structural covariance invariance]; *p<.001

3.6.2 Development and standardization of perceived teacher engagement scale (PTES)

Work engagement research is gaining popularity, however there is an ongoing dispute over how to define and assess it (Bakker, Albrecht, & Leiter, 2011; Jeung, 2011; Shuck, Ghosh, Zigarmi, & Nimon, 2012). According to Schaufeli et al. (2002), work engagement is regarded as a psychological state as opposed to burnout. Unlike teachers who are burnt out, engaged teachers feel energetic and establish strong connection with their work activities. Over the years, the three-dimensional measure of work engagement established by Schaufeli et al., (2002) has emerged as the most popular instrument (Shuck, 2011; Schaufeli & Salanova, 2011*).* Initially, Schaufeli and Bakker (2003) introduced the UWES-17, a seventeen-item questionnaire, and subsequently Schaufeli et al. (2006) offered a nine-item shorter version of the instrument (UWES-9). The tool has been translated and validated to measure the work engagement construct across various cultural contexts. For example, the Japanese version (Shimazu et al., 2008), the Italian version (Balducci, Fraccaroli, & Schaufeli, 2010), the Chinese version (Yi-wen, & Yi-qun, 2005), the Russian version (Lovakov et al., 2017), the French version (Zecca et al, 2015) etc. have been widely used.

In sum, it can be noted from the early studies that there are only a few self-report instruments to measure teacher engagement. Among those, the UWES (Schaufeli et al.,

2002) is mostly used across studies. Besides, another instrument namely, the Engaged teacher scale (ETS) developed by Klassen, Yerdelen, & Durksen (2013) is also used by many scholars for measuring self-reported engagement of the teachers. ETS is a seven-point Likert scale with 16 items following a 4-factor measurement model (see Table 3.26) of teacher engagement. However, how students perceive teacher engagement has a high impact on their classroom behavior and on positive learning outcomes as well. For example, students who perceive that teachers' higher engagement in teaching are likely to be motivated in learning. Thus, student perceptions of teacher engagement may result in several positive student outcomes including formation of positive classroom environment, students' self-confidence and their active involvement in classroom learning. Besides, students' feelings of teacher engagement may relate to positive classroom behavior of the students, student engagement in learning activities, and creating a supportive learning environment.

However, all the available instruments (see Table 3.26) measure teachers' engagement where teachers are the respondents. Thus, the researchers have identified the gap in the literature that there is no instrument available to gauge the students' perceptions of teacher engagement. To fill the gap, we focused on the development of a research tool regarding students' perceptions of teacher engagement. The study hence aimed to develop and validate perceived teacher engagement scale (PTES) and evaluate the psychometric properties of PTES. Additionally, it is crucial to develop a 'strong' (i.e. equal factor loadings and equal intercepts; Byrne, 2008) measurement invariance of the measures in order to compare the groups on the basis of latent mean scores. The degree to which the measuring instrument's parameters are comparable across groups is known as measurement invariance (Byrne, 2008). Measurement invariance is generally examined using multi-group CFA. In the analysis it is assumed that the factorial structure of the same latent construct(s) remains identical across all the groups (Byrne & van de Vijver, 2010). However, if an assessment instrument's equivalence or invariance is not maintained, the inferences drawn from the analysis of data collected using the instrument may be invalid (Byrne, 2008), and the conclusions based on group comparisons would be erroneous. In this study, the test of measurement invariance of the instrument was performed across gender. If the instrument qualifies the gender-invariance then it can be

employed to compare the perceptions of teacher engagement across male and female groups of students.

3.6.2.1 Item writing

From the reviews of articles on teacher/work engagement (Schaufeli et al., 2006; Yerdelen et al., 2018) the following seven dimensions of teacher engagement were identified as common attributes irrespective of the instruments: (1) cognitive engagement, (2) emotional engagement, (3) social engagement: colleagues, (4) social engagement: students, (5) vigor, (6) dedication, (7) absorption. The definitions of each of these dimensions are presented in Table 3.26. The items for PTES were developed on the basis of those dimensions and items keeping the students' perspectives in the conscience. An initial pool of 87 items was developed by following the nature of the items under each factor while considering the specific generic essence of each of the factors.

3.6.2.2 Initial Try-out:

A planned iterative process was followed to exclude the items that merely served as antecedents and consequences of perceived teacher engagement construct. The initial try-out was conducted to ensure that items present in the tools covered all the areas of the construct, and the items are consistent with the context of the present study.

A. Expert Try-out:

The preliminary draft of the target instrument containing 87 scale items was sent to seven experts (Lambie et al., 2017) for their suggestions and recommendations. The scholars were selected as experts whose area of specialization is measurement and evaluation in Psychology and Education and who have minimum 10 years of experience in teaching and research in those specified fields (Ikart, 2019). In view of the criticism and comment by the experts, 27 items were removed from the Perceived Teacher Engagement Scale (PTES) while three were modified accordingly. Thus, the final draft with 60 items (33 negative items) was prepared considering feedback from the experts. Then, this draft was also approved by the Departmental Research Committee (DRC), Department of Education, Tezpur University. This allowed the researcher to conduct the pre-test (individual try-out) of the tool.

Table 3.26

Major dimensions, their definition/characteristics, sample items and their related sources.

Instrument used	Factors of teacher engagement	Definition and/or characteristics	Sample items	Related sources
Utrecht work engagement scale (UWES)	F1:Vigor	F1: "Vigor is characterized by high levels of energy and mental resilience while working, the willingness to invest effort in one's work, and persistence even in the face of difficulties" [Schaufeli et al., (2002), p. 74-75].	F1: 'At my work, I feel that I am bursting with energy'.	Schaufeli et al. (2002; 2006), Han, Perron, Yin, & Liu (2020); Abiodullah, Sameen, & Aslam, (2020); Kulophas, Hallinger, Ruengtrakul, & Wongwanich (2018); Li, Liu, Chen, & Yao (2019); Yin, Han, & Lu (2017)
	F2:Dedication	F2: "Dedication is characterized by a sense of significance, enthusiasm, inspiration, pride, and challenge" [Schaufeli et al., (2002), p. 74–75].	F2: 'I feel happy when I am working intensely'.	
	F3:Absorption	F3: "Absorption is characterized by being fully concentrated and deeply engrossed in one's work, such time passes quickly and one has difficulties in detaching oneself from work" [Schaufeli et al., (2002), p. 74–75].	F3: 'At my job, I am very resilient, mentally'.	

Engaged teacher scale (ETS)	F1:Emotional engagement	F1: Positive feelings, satisfaction and attachment to the profession	F1: 'I really put my heart into teaching'.	Yerdelen, Durksen, & Klassen (2018); Klassen, Yerdelen, & Durksen (2013); Sokmen & Kilic (2019); Karahan (2018)
	F2:Cognitive engagement	F2: The extent of concentration and persistence in work activities	F2: 'While teaching, I get absorbed in my work'.	
	F3:Social engagement: students	F3: Efforts, skills and activities to build rapport with students	F3: 'I connect well with my students'.	
	F4:Social engagement: colleagues	F4: Efforts, skills and activities to build rapport with colleagues.	F4: 'I am accessible to my colleagues'.	

B. Individual Try-out (Pre-test and Pilot study):

After Experts Tryout, the student engagement in learning scale was left with 60 items. To reduce measurement error (Blair & Conrad, 2011), pre-testing was conducted with these 60 scale items on 30 students (Perneger, Courvoisier, Hudelson, & Gayet-Ageron, 2015) of Class-X who were not considered in the final study sample (Cooper and Schindler, 2011; Kumar, Talib & Ramayah, 2013). This pre-testing was performed to check whether any of the items was ambiguous, lacks in clarity, not appropriately sequenced in the scale, and whether the instructions were given properly to the respondents (Kumar et al., 2013).

Results of the pre-testing suggested that for 13 items the scores (e.g. Teacher treats students getting good grades better than others, Teacher does not explain what we are going to learn in the class) showed minor deviations and low skewness from the mean score. Those items were excluded from the scale. Additionally, 4 items (e.g. Teacher helps me with my problems, Teacher does not explain what we are going to learn in the class, Teacher motivates us to read books other than prescribed textbook) were deleted due to their ambiguity and lack of clarity to the respondents, whereas 6 items were dropped (e.g. Teacher reads out from textbook in class without explaining anything) owing to item social desirability bias (Memon, Ting, Ramayah, Chuah, & Cheah, 2017). In sum, on the basis of individual try-out, 23 items were removed from the preliminary draft and 12 items were slightly modified. Thus, the preliminary draft of the SELS included 37 items (18 negative items).

Further, the revised questionnaire with 37 items underwent a second round of pre-testing (Memon et al., 2017) that suggested for no further modifications. Finally, a rating scale with 37 items (18 negative) was developed which was to be rated by the students based on their perceptions. Thus, the modification in the items was found to be effective. Finally, the instrument with 37 items was piloted on another 30 students (Memon et al., 2017). The scores were reversed for the negative items before conducting any calculation. Then, a reliability analysis was performed considering all the items representing a single construct. The results depicted acceptable reliability as the Cronbach alpha (α) value was

0.92. However, the factorial structure of the construct was further tested in Group Try-out phase.

C. Group Try-out:

The Perceived Teacher Engagement Scale (PTES) with 37 items was administered to a group of 392 tenth graders in government secondary schools in Birbhum, West Bengal. Among them 188 (47.959%) were male and 204 (52.041%) were female. Consents were taken from the head of the institutions for getting the responses from the students. Then, each student was provided with one copy of the student engagement in learning scale and was asked to follow the instructions. Further, they were informed that their responses will neither be evaluated nor be disclosed anywhere.

The items were assigned with the scores from '5' to '1' for 'Strongly Agree', to 'Strongly Disagree', respectively. However, reverse scoring procedure was applied for the negative items. The total score of each respondent was calculated by adding the scores in each engagement sub-scale. Lesser values of total scores indicated lower level of engagement whereas the higher values of the total scores indicated higher level of engagement.

Further, item analysis was performed for 37 scale items. The raw scores of the items those were obtained in the form of the responses from 392 students were used for item analysis. Firstly, from the responses obtained from Group Tryout, individual scores were calculated and were arranged in descending order. Then, respondents belonging to the top 27% group (N= 106) and bottom 27% (N= 106) group were identified. Then, the data of 106 subjects belonging to the top 27% group and 106 belonging to the bottom 27% group were analyzed for 37 scale items to calculate the Discrimination Index of the items (Table 3.27):

Table 3.27

Item-wise M, SD, and t-values

N= 212

Items	M_1	SD_1	M_2	SD_2	t	p-value	Remark
1	3.14	1.03	3.34	0.68	-1.69	0.09	Rejected
2	3.21	0.86	3.34	0.93	-1.04	0.30	Rejected
3	3.30	0.89	3.64	0.95	-2.76	0.01	Selected

4	3.15	0.84	3.18	0.84	-0.24	0.81	Rejected
5	3.22	0.91	3.12	0.86	0.91	0.36	Rejected
6	3.12	0.82	3.13	0.87	-0.08	0.94	Rejected
7	3.21	0.81	3.16	0.90	0.39	0.70	Rejected
8	3.10	0.80	3.18	0.88	-0.71	0.48	Rejected
9	3.30	0.90	3.16	0.83	1.23	0.22	Rejected
10	3.31	0.92	3.11	0.83	1.75	0.08	Rejected
11	3.20	0.92	3.12	1.12	0.59	0.56	Rejected
12	3.14	0.76	3.13	0.89	0.08	0.94	Rejected
13	3.37	0.90	2.90	0.71	4.29	0.00	Selected
14	3.30	0.90	2.81	0.90	4.10	0.00	Selected
15	3.46	0.82	3.22	0.95	2.04	0.04	Selected
16	3.39	0.88	3.08	1.05	2.41	0.02	Selected
17	3.47	0.93	3.28	1.02	1.51	0.13	Rejected
18	3.50	0.96	3.07	1.04	3.21	0.00	Selected
19	2.84	1.17	3.34	0.68	-3.92	0.00	Selected
20	2.92	1.01	3.26	0.79	-2.81	0.01	Selected
21	3.09	0.98	3.27	0.72	-1.55	0.12	Rejected
22	3.31	1.02	3.23	0.66	0.70	0.48	Rejected
23	3.25	0.92	2.68	0.76	5.08	0.00	Selected
24	2.81	1.38	3.35	1.14	-3.17	0.00	Selected
25	3.25	0.80	2.86	0.91	3.43	0.00	Selected
26	3.19	0.81	2.71	0.91	4.19	0.00	Selected
27	3.38	0.88	2.79	0.88	5.00	0.00	Selected
28	2.24	1.36	3.09	0.82	-5.66	0.00	Selected
29	2.74	1.30	3.11	1.13	-2.25	0.03	Selected
30	3.24	0.82	2.67	1.13	4.32	0.00	Selected
31	3.21	0.74	2.62	0.95	5.25	0.00	Selected
32	2.63	1.27	3.18	1.11	-3.47	0.00	Selected
33	2.88	0.94	3.38	1.02	-3.81	0.00	Selected
34	2.95	0.98	3.04	1.04	-0.72	0.47	Rejected
35	3.04	0.72	2.58	1.21	3.41	0.00	Selected
36	3.10	0.86	2.77	0.81	2.97	0.00	Selected
37	3.05	0.85	2.75	1.14	2.26	0.02	Selected

From the Table 3.27, it is clear that t-values for the items at serial numbers 3, 13, 14, 15, 16, 18, 19, 20, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37 are significant at least at 0.05 level. These items were found to discriminate between the respondents belonging to the top 27% group and the bottom 27% group. Thus, these 22 scale items were kept in the instrument for further analysis.

3.6.2.3 Final Try-out:

Data were collected from two independent randomized sub-samples for conducting EFA and CFA. The first sub-sample contained 563 (276 females) and second sub-sample contained 579 (286 females) secondary school students in the seven districts (Cooch Behar, Purba Bardhaman, Nadia, Murshidabad, Birbhum, North Dinajpur and Hooghly) of West Bengal under WBBSE. None of the samples was included in the final data collection of the study. Bengali was the medium of instruction and communication in all the selected schools. Consents were taken from both the head of the institutions and of the students for getting the responses. Then, each student was provided with one copy of the PTES and was asked to follow the instructions. Further, they were informed about the confidentiality of the responses.

A. Preliminary descriptive analyses

Before doing the main analyses, data were examined to check whether the statistical assumptions were fulfilled. There was no missing value in the dataset. Measures of central tendency and variability were performed for each scale item (see Table 3.28). Results (see Table 3.28) showed that the measures were within the statistically acceptable range for all items (skewness<2 and kurtosis<7; Curran, West, & Finch, 1996).

The means and SDs ranged from 2.58 to 4.32 and from 0.98 to 1.16, respectively. The standardized value (Z) of the skewness and kurtosis (see Table 3.28) for all items were within the statistical thresholds (Kline, 2011). Further, the item total correlations and the inter-item correlations for each scale item were calculated and the results have been demonstrated in Table 3.29 and in Table 3.30, respectively.

Table 3.28

Mean, SD, skewness and kurtosis of the 22 scale items

	Mean	Std. Deviation	Skewness		Kurtosis	
Item no.	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
1	3.31	1.037	-0.226	0.103	-0.351	0.205
2	3.24	1.096	-0.258	0.117	-0.513	0.208
3	4.32	1.063	-0.257	0.100	-0.424	0.207
4	3.36	1.103	-0.309	0.104	-0.476	0.202
5	3.28	1.074	-0.675	0.088	-0.190	0.213
6	3.52	1.073	-0.710	0.091	-0.183	0.163
7	3.49	1.073	-0.364	0.104	-0.338	0.152
8	3.13	1.080	-0.017	0.108	-0.520	0.171
9	3.11	1.079	-0.053	0.084	-0.614	0.168
10	3.10	1.083	-0.179	0.087	-0.469	0.174
11	3.32	1.073	-0.168	0.127	-0.408	0.197
12	3.21	1.076	-0.125	0.124	-0.475	0.209
13	3.25	1.054	-0.385	0.109	-0.424	0.201
14	3.27	1.053	-0.319	0.107	-0.511	0.200
15	3.29	1.103	-0.235	0.088	-0.517	0.207
16	3.23	1.073	-0.309	0.115	-0.411	0.251
17	3.20	1.063	-0.326	0.114	-0.441	0.201
18	3.25	1.101	-0.311	0.081	-0.505	0.121
19	3.37	1.136	-0.615	0.101	-0.398	0.196
20	2.97	1.160	-0.603	0.107	-0.514	0.203
21	2.58	1.141	-0.494	0.106	-0.547	0.212
22	2.87	1.146	-0.368	0.101	-0.638	0.211

Table 3.29

Item-Total Statistics

Items	Corrected Item-Total Correlation	Items	Corrected Item-Total Correlation
1	0.42	12	0.52
2	0.40	13	0.51
3	0.41	14	0.52
4	0.45	15	0.42
5	0.43	16	0.46
6	0.38	17	0.47
7	0.38	18	0.41

8	0.45	19	0.43
9	0.44	20	0.47
10	0.51	21	0.46
11	0.50	22	0.45

A. Main analyses

Data analysis was performed using SPSS 26.0 and AMOS 23.0 module. Firstly, to develop the measurement model for student engagement EFA was conducted with the 22 scale items. Reverse scoring was carefully done for the negative items before performing any analysis.

Next, to test and evaluate the measurement model developed using EFA, a series of Confirmatory Factor Analyses (CFA) was performed. Firstly, a zero-order 1-factor model was evaluated by taking the entire set of 22 items together and loaded on a single factor. Next, three 2-factor models were tested taking anyone out of the three factors (viz. cognitive, behavioral, and emotional engagement) as a distinct factor and combining the items of the rest of the two factors. Finally, a first-order 3-factor model was evaluated. The 5 measurement models were then compared on the basis following model fit indices: $\chi 2$ statistic and associated p-value, CFI, GFI, PCFI, PNFI, RMSEA, and SRMR, where the threshold values for CFI and GFI was ≥.90 (Bentler & Bonnet, 1980), ≥.50 for PCFI and PNFI (James, Mulaik, & Brett, 1982), and <.80 for acceptable (MacCallum, Browne, & Sugawara, 1996) or ≤.60 (Hu & Bentler, 1999; Vanderberg & Lance, 2000) for RMSEA and SRMR for a good fit.

Besides, the mean scores of student engagement dimensions were compared to evaluate whether the evaluation of the factors differs across gender of the students. The factorial invariance of the instrument was evaluated using multi-group confirmatory factor analysis (MGCFA) (Byrne, 2010). Finally, inter-correlations among the dimensions of the construct were also calculated using Pearson's product-moment correlations.

Table 3.30

Inter-item Correlations for PTES

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1																					
2	0.21	1																				
3	0.24	0.21	1																			
4	0.37	0.32	0.13	1																		
5	0.22	0.36	0.06	0.23	1																	
6	0.19	0.04	0.24	0.22	0.28	1																
7	0.16	0.18	0.19	0.18	0.27	0.34	1															
8	0.03	-0.04	-0.05	-0.01	-0.04	-0.03	-0.08	1														
9	0.05	0.03	0.04	0.06	0.03	-0.03	-0.02	0.19	1													
10	0.08	0.01	0.03	0.06	0.05	0.03	0.01	0.17	0.12	1												
11	0.02	0.03	0.05	0.07	0.03	0.01	-0.01	0.12	0.15	0.21	1											
12	0.05	0.03	0.04	0.09	0.04	0.04	-0.01	0.11	0.24	0.28	0.17	1										
13	0.08	0.09	0.08	0.11	0.09	0.09	0.07	0.12	0.21	0.31	0.29	0.24	1									
14	0.07	0.10	0.05	0.10	0.04	0.06	0.04	0.31	0.33	0.17	0.20	0.27	0.23	1								
15	0.03	0.01	-0.02	0.04	0.00	-0.01	-0.06	0.34	0.28	0.13	0.21	0.25	0.11	0.13	1							
16	0.09	0.12	0.10	0.08	0.09	0.05	0.10	0.22	0.08	0.13	0.13	0.09	0.12	0.14	0.06	1						
17	0.08	0.06	0.15	0.14	0.14	0.06	0.07	0.12	0.08	0.15	0.15	0.11	0.11	0.16	0.10	0.29	1					
18	0.11	0.06	0.12	0.10	0.14	0.09	0.10	0.06	-0.02	0.10	0.08	0.07	0.07	0.08	0.04	0.17	0.25	1				
19	0.06	0.09	0.14	0.11	0.09	0.09	0.13	0.09	0.02	0.10	0.13	0.08	0.03	0.12	0.03	0.08	0.32	0.13	1			
20	0.09	0.08	0.10	0.14	0.13	0.07	0.11	0.06	0.02	0.10	0.09	0.10	0.11	0.13	0.09	0.02	0.19	0.27	0.28	1		
21	0.08	0.10	0.11	0.12	0.11	0.10	0.13	0.03	0.03	0.10	0.11	0.10	0.06	0.12	0.06	0.07	0.27	0.36	0.11	0.23	1	
22	0.08	0.08	0.11	0.13	0.13	0.08	0.10	0.05	0.02	0.11	0.10	0.07	0.07	0.11	0.00	0.11	0.18	0.28	0.13	0.25	0.17	1

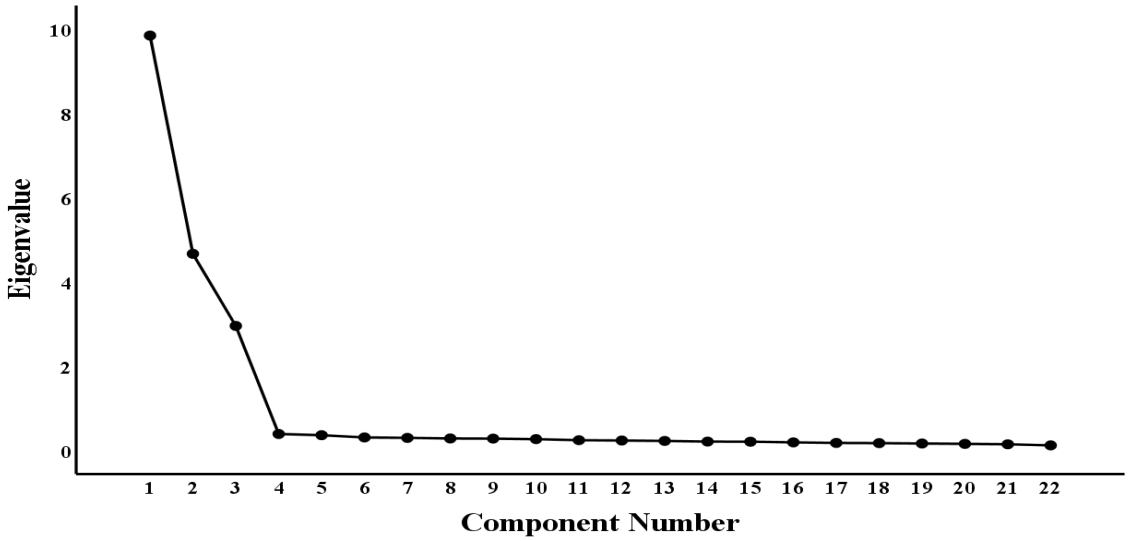
I. Development of the measurement model: Exploratory Factor Analysis

To explore the factors of student engagement, EFA was performed using promax rotation and principal components analysis as the method of extraction of factors. The oblique rotation used to allow that the factors of the construct are related (Byrne, 2008). Before that, the prerequisites for conducting EFA were checked. Firstly, Kaiser-Meyer-Olkin's measure of sample adequacy (.963) was greater than 0.6 and Bartlett's Test of Sphericity was significant ($\chi 2$=12835.725, df=231, p<.001). Further, the non-zero determinant value ($8.604 \times e^{-11}$) of the correlation matrix satisfied the criteria of positive definiteness.

Results of the EFA came up with three-factor solution following Kaiser Criterion (eigenvalue>1; Cattell, 1966). The examination of the scree plot was supported the three-factor solution (see Graph 3.3). Further, to determine the number of factors to be retained in the measurement model, Parallel analysis (O'Connor, 2000) was run. The comparison between the eigenvalues of raw data from EFA with 95^{th} percentile of random eigenvalues suggested retaining three factors (see Graph 3.4). This can be visualized in Figure 3.9. Additionally, Wayne Velicer's Minimum Average Partial (MAP) test (Velicer, 1976; Velicer et al., 2000) indicated that the three-factor structure of perceived teacher engagement construct achieved both a minimum of 0.01 average partial and the lowest Bayesian information criterion (BIC) as suggested from the parallel analysis.

Graph 3.3

Scree plot for the 3-factor solution of EFA



Graph 3.4

Tsplot of Raw Data Eigenvalues, & Mean & (95th) Percentile Random Data Eigenvalues

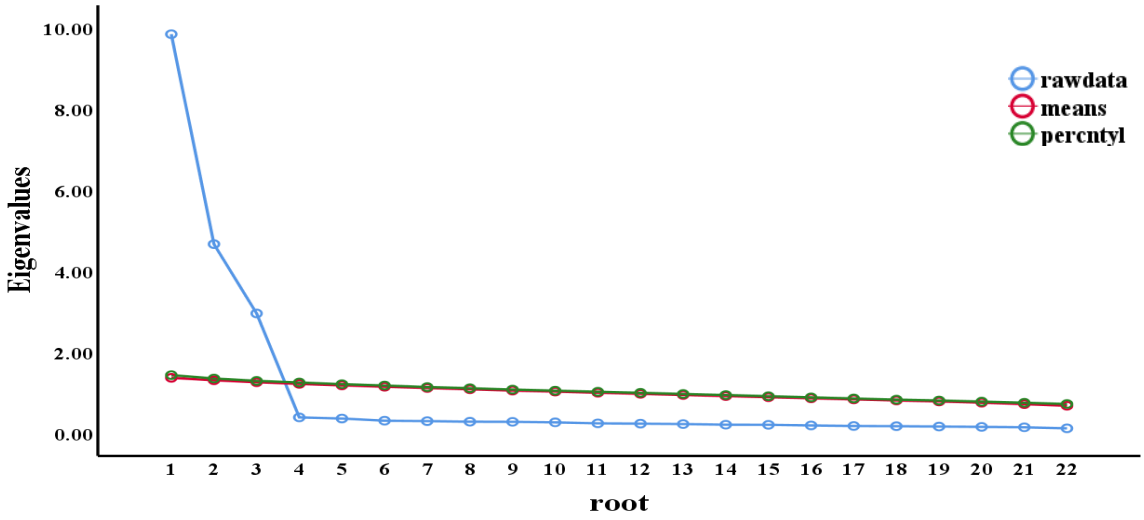


Table 3.31

Raw Data Eigenvalues, & Mean & (95th) Percentile Random Data Eigenvalues

Root	Raw Data Eigenvalues	Means of Eigenvalues	(95th) Percentile Eigenvalues
1	9.842	1.370	1.437
2	4.667	1.310	1.354
3	2.958	1.262	1.295
4	0.395	1.221	1.256
5	0.367	1.184	1.214
6	0.313	1.151	1.181
7	0.303	1.120	1.142
8	0.288	1.088	1.114
9	0.286	1.056	1.081
10	0.273	1.031	1.054
11	0.248	1.003	1.027
12	0.241	0.976	1.000
13	0.230	0.947	0.972
14	0.215	0.920	0.946
15	0.212	0.893	0.917
16	0.196	0.866	0.888
17	0.183	0.840	0.862
18	0.178	0.812	0.833
19	0.169	0.788	0.811
20	0.161	0.758	0.784
21	0.151	0.723	0.755
22	0.125	0.681	0.725

The measurement model explained 79.851% of the variance in the latent construct in total. The rotated component matrix showed that the standardized factor loadings for all the items were greater than |.45| (Hair, Tatham, Anderson, and Black, 1998) (see Table 3.32). The items were assigned to each scale dimension following their loading patterns.

The nomenclature of the dimensions was based on the overall nature of the included items for each corresponding dimensions. The three-factor model was developed: Perceived Cognitive-Physical Engagement (PCPE) (item no.: 1 to 7 [PCPE1 to PCPE7]), Perceived Socio-Emotional Engagement (PSEE) (item no.: 8 to 15 [PSEE1 to PSEE8]), Perceived Pedagogical Engagement (PPE) (item no.: 16 to 22 [PPE1 to PPE7]) (see Table 3.32).

The first subscale i.e. PCPE accounted for 29.611% of the total variance in the construct. This factor included items such as “I ask the teacher for clarification” and “I remain attentive in the class”. The second subscale (PSEE) accounted for 26.845% of the total variance in student engagement. Some items are: “Performing well in school is important to me” and “I plan to continue for higher studies”. Finally, the third subscale (PPE) accounted for 23.395% of the variance. Some of the items are: “I feel regretted when I miss any class” and “The classes are boring”. Finally, Cronbach’s alpha values were found to be 0.967 for cognitive, 0.962 for behavioral, 0.959 for emotional engagement, and 0.912 for the overall scale that indicates high reliability of the measurement scale along with its dimensions.

To examine whether the dimensions of student engagement are significantly related to each other, Correlational analyses (Pearson’s product-moment correlation) were performed. However, all bivariate correlation coefficients were positive and statistically significant. The results showed that perceived cognitive-physical engagement showed slightly higher correlation with perceived socio-emotional engagement dimension ($r = 0.223$, $p < 0.01$, effect size= 0.050) than with perceived pedagogical engagement dimension ($r = 0.370$, $p < 0.01$, effect size= 0.137) (see Table 3.32).

Table 3.32

Rotated Component Matrix: Standardized factor loadings of 22 items in three dimensions of the Perceived Teacher Engagement construct

	Components		
Scale items	**1**	**2**	**3**
Perceived Cognitive-physical engagement items			
1. Teacher puts all effort to make us understand lessons.	.905		
2. Teacher remains seated on the chair during the entire class hour.*	.893		
3. Teacher gets irritated when I ask questions.*	.895		
4. Teacher never gets tired of teaching.	.911		
5. Teacher gets diverted during teaching in class.*	.918		
6. Teacher takes phone calls while teaching in class.*	.886		
7. Teacher comes to class with preparation.	.875		
Perceived Socio-emotional engagement items			
8. Teacher does not care whether I learn.*		.898	
9. Teacher believes that my performance can be better.		.885	
10. Teacher calls students by their names.		.906	
11. Teacher does not listen to the students.*		.905	
12. Teacher follows whether I am doing my tasks in class.		.936	
13. Teacher is available for discussions on studies beyond class hours.		.884	
14. Teacher always blames students for their failure.*		.865	
15. Teacher talks interestingly in class.		.858	
Perceived Pedagogical engagement items			
16. Teacher provides interesting activities in class.			.876
17. Teacher never asks questions on previously taught lessons.*			.869
18. Teacher does not use teaching aids in class.*			.869
19. Teacher rarely use blackboard in class.*			.863
20. Teacher connects classroom discussions to daily life experiences.			.902
21. Teacher does not encourage us to ask questions.*			.888
22. Teacher makes clear why my answers are good or not.			.898
Factor inter-correlations			
Factor 1	-	.223**	.370**
Factor 2		-	.311**
Factor 3			-

Note. Correlation coefficients ≤0.45 are not shown; *Negative scale items, **p<.001

Further, the correlation between perceived socio-emotional engagement and perceived pedagogical engagement was ($r = 0.311$, $p < 0.01$, effect size= 0.097) low but significant (see Table 3.32). This shows that the dimensions are interrelated among each other. However, the low ($r \leq .10$) to medium values ($.10 \leq r \leq .30$) of effect size (Cohen, 1988) signify that those intra-construct correlations were practically small in size. This signifies that even the factors are highly interrelated to each other, those measure different aspects of the student engagement construct. Hence, those are considered to be three distinct constructs.

The final draft of Perceived Teacher Engagement Scale consisted with 22 items (11 negative). The dimension-wise distribution of serial number-wise items has been depicted in Table 3.33.

Table 3.33

Serial Number-wise distribution of items in each dimension and types of items

Sl. No.	Dimensions	Nature of items	Sl. No. of items in Final Scale	Total no. of items	Total
1	Perceived Cognitive-Physical Engagement	Positive	1, 4, 7	3	7
		Negative	2, 3, 5, 6	4	
2	Perceived Socio-Emotional Engagement	Positive	9, 10, 12, 13, 15	5	8
		Negative	8, 11, 14	3	
3	Perceived Pedagogical Engagement	Positive	16, 20, 22	3	7
		Negative	17, 18, 19, 21	4	
	Positive items= 11 + Negative items= 11			Total items	22

II. Confirmation of the measurement model: Confirmatory Factor Analysis

Results of the zero-order CFA considering all 22 items as a single-factor (1-factor model) showed poor model-fit indices: $\chi^2 = 7955.452$, $p < 0.001$, $\chi 2/df = 38.064$, CFI = 0.396, GFI= .312, PCFI= .358, PNFI= .353, RMSEA = 0.253, and SRMR = 0.366. Similarly, none of the competing 2-factor models were supported for poor model fit resulted from a series of first-order CFA (see Table 3.34). Finally, a first-order CFA (Byrne, 2005) with

the final 22 scale items under the three factors (PCPE, PSEE and PPE) of student engagement was conducted. The model fit indices satisfactorily supported the 3-factor model: $\chi 2$= 336.519, p<0.001, $\chi 2$/df = 1.634, CFI= 0.990, GFI= .949, PCFI= .883, PNFI= .869, RMSEA= 0.033, SRMR = 0.030. Therefore, the three-factor model outperformed both the single-factor as well as the two-factor models. Finally, the 3-factor model of perceived teacher engagement (see Figure 3.6) was retained as it showed better fit compared to other competing 1-factor model and 2-factor models. The standardized factor loadings are given in Figure 3.35.

Table 3.34

Model fit indices of 1-factor, 2-factor, and 3-factor models from confirmatory factor analyses

	χ^2	p	χ^2/df	CFI	GFI	PCFI	PNFI	RMSEA	SRMR
One factor model									
Model 1: PCPE/PSEE/PPE[a]	7955.45	<.001	38.06	.396	.312	.358	.353	.253	.366
Two factor models									
Model 2: PCPE+PSEE/PPE	3981.97	<.001	19.14	.706	.481	.635	.626	.177	.270
Model 3: PSEE+PCPE/PPE	4479.24	<.001	21.54	.667	.453	.600	.591	.188	.280
Model 4: PPE+PCPE/PSEE	4314.36	<.001	20.74	.680	.460	.612	.603	.185	.251
Three factor model									
Model 5: PCPE+PSEE+PPE	336.519	<.001	1.63	.990	.949	.883	.869	.033	.030

[a]*Note.* A single engagement scale (either PCPE or PSEE or PPE) refers to a distinct factor with the items solely loaded on that factor. A set of engagement scales that appears in a group (e.g. PCPE/PSEE/PPE) refers to a common combined factor with all items from those factors loaded on that single composite factor. The + sign was used to mean distinct factors and the / sign was used to mean the combination of items in two or more factors

Figure 3.6

Three-factor measurement model of perceived teacher engagement

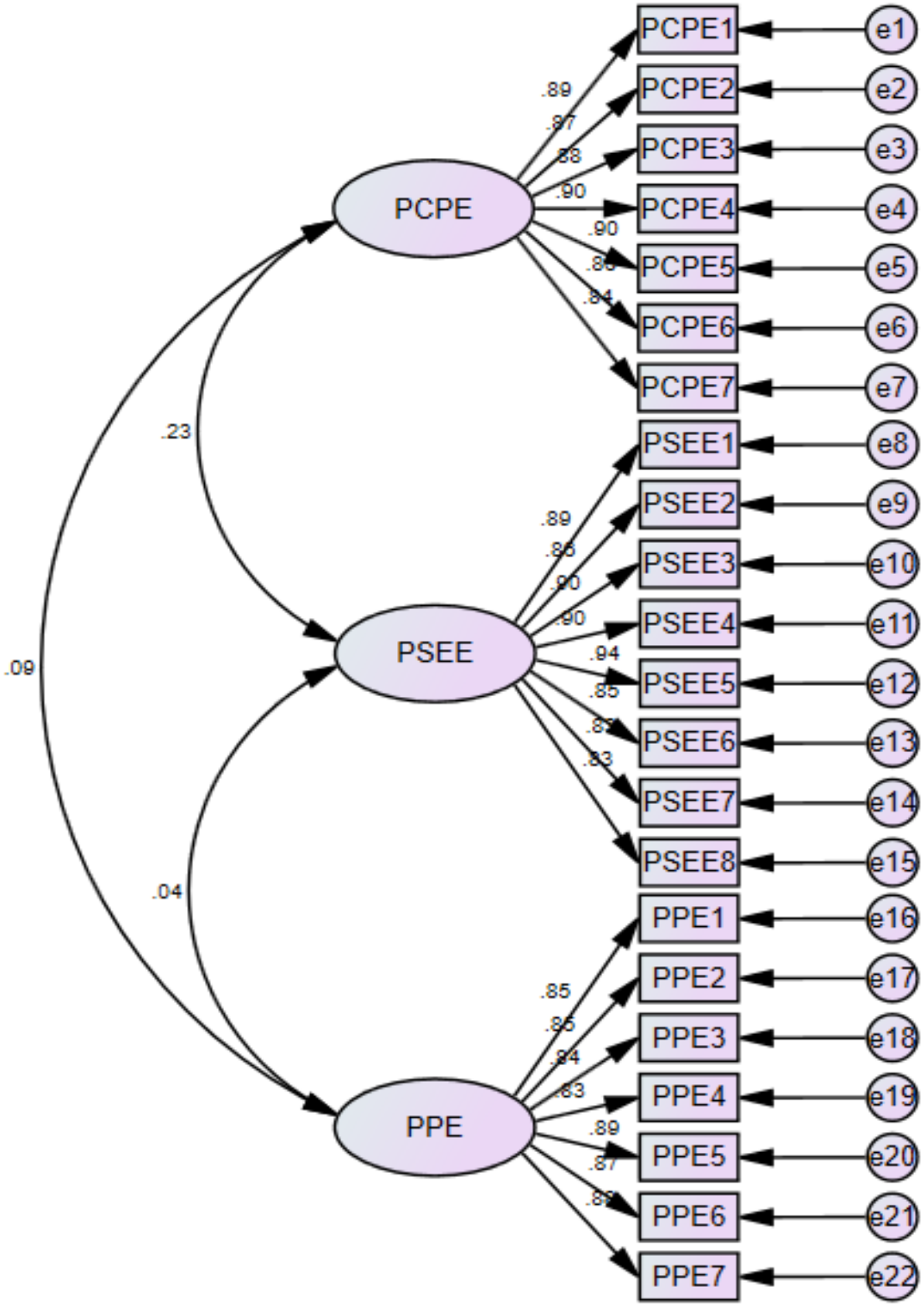


Table 3.35

Standardized Regression Weights

Regression path	Estimate	Regression path	Estimate	Regression path	Estimate
PCPE1<--- PCPE	.892	PSEE1<--- PSEE	.885	PPE1<--- PPE	.851
PCPE2<--- PCPE	.874	PSEE2<--- PSEE	.862	PPE2<--- PPE	.847
PCPE3<--- PCPE	.875	PSEE3<--- PSEE	.896	PPE3<--- PPE	.842
PCPE4<--- PCPE	.899	PSEE4<--- PSEE	.897	PPE4<--- PPE	.826
PCPE5<--- PCPE	.905	PSEE5<--- PSEE	.936	PPE5<--- PPE	.886
PCPE6<--- PCPE	.855	PSEE6<--- PSEE	.852	PPE6<--- PPE	.868
PCPE7<--- PCPE	.843	PSEE7<--- PSEE	.831	PPE7<--- PPE	.879
		PSEE8<--- PSEE	.827		

III. Evaluation of the psychometric properties of PTES

a. Validity

The scale has Face Validity as the scale was examined by the experts and the scale was modified accordingly based on their suggestions. For content validity, item analysis was conducted and only the items demonstrating statistically significant discriminative power were retained. Further, the Factorial Validity of the scale was also examined using First-order CFA. The results are given in Table 3.36 and Table 3.37.

i. Convergent validity:

The relationship between the observed variables and their corresponding latent variables was examined using convergent validity where the standardized factor loadings must be ≥0.55 (Hair, Hult, Ringle, & Sarstedt, 2017). In the three-factor model all the standardized factor loadings were statistically significant ($p<0.001$) and ranged between 0.826 and 0.936. Further, the composite reliability (CR) coefficients for each dimension of student engagement (perceived cognitive-physical, perceived socio-emotional, and perceived pedagogical engagement was 0.959, 0.963, and 0.926, respectively; see Table 3.36) demonstrated a satisfactory value above 0.7 (Fornell, 1982). Whereas, the average variance extracted (AVE) for each dimension (perceived cognitive-physical, perceived socio-emotional, and perceived pedagogical engagement was 0.771, 0.764, and 0.643, respectively; see Table 3.36) was greater than 0.50 (Fornell, 1982) which indicated that a greater common variance was captured by each construct than the variance due to the

measurement error. These results indicate that each of the dimensions qualifies for convergent validity (CR>.7, AVE>.5, CR>AVE; Hair et al., 2017).

Table 3.36

Convergent validity of the engagement sub-scales

Sub-scales	CR	AVE	CR>0.7	AVE>0.5	CR>AVE	Convergent Validity
PCPE	0.959	0.771	Satisfied	Satisfied	Satisfied	Established
PSEE	0.963	0.764	Satisfied	Satisfied	Satisfied	Established
PPE	0.926	0.643	Satisfied	Satisfied	Satisfied	Established

ii. Discriminant validity:

Further, maximum shared variance (MSV) (for PCPE=0.053, PSEE=0.053 and PPE=0.008; see Table 3.37) and average shared variance (ASV) (for PCPE=0.026, PSEE=0.018 and PPE=0.004; see Table 3.37) for each latent construct were calculated from intra-construct correlation coefficients in the 3-factor measurement model. Both MSV and ASV were found to be numerically less than AVE for each latent factor. Hence, all the factors qualify for discriminant validity (Hair, Black, Babin, Anderson, & Tatham, 2014).

Table 3.37

Discriminant validity of the engagement sub-scales

Sub-scales	AVE	MSV	ASV	AVE>MSV	AVE>ASV	Discriminant Validity
PCPE	0.771	0.053	0.026	Satisfied	Satisfied	Established
PSEE	0.764	0.053	0.018	Satisfied	Satisfied	Established
PPE	0.643	0.008	0.004	Satisfied	Satisfied	Established

b. Reliability

The internal consistency of the Student Engagement in Learning Scale was calculated using Cronbach's Alpha for each sub-scale and for the overall scale based on the scores of 579 respondents. The reliability coefficients are given in Table 3.38. The Cronbach's α value (Cronbach. 1951) for PCPE, PSEE, and PPE was 0.910, 0.932, and 0.911, respectively, and 0.873 for the overall scale was higher than the statistically acceptable figures (α>.7; Hair et al., 2017; see Table 3.38). This showed the high reliability of the instrument. Besides, the Split-Half Reliability of the scale was obtained employing odd-even method on the scores of 579 students. The scores of the students on odd numbered items and on even numbered items were put into the Spearman-Brown Prophecy formula to obtain the Split-Half Reliability Coefficient of 0.91 for the Student Engagement in Learning Scale, which is significant at 0.01 level.

Table 3.38

Reliability coefficient of the overall scale and dimensions of Perceived Teacher Engagement in Learning Scale

Student Engagement Dimensions	Number of items	Cronbach's α
1. Perceived Cognitive-Physical Engagement	7	0.910
2. Perceived Socio-Emotional Engagement	8	0.932
3. Perceived Pedagogical Engagement	7	0.911
Overall Perceived Teacher Engagement Scale	22	0.873

Therefore, the results of CFA confirmed the construct validity and internal consistency reliability of the three-factor measurement model of student engagement as obtained from the results of the EFA. The present study upholds the factorial structure of the perceived teacher engagement construct as proposed by the earlier researchers and confirms the absence of the fourth factor as advocated by other researchers.

c. Norms

In Final Tryout, the reliability and validity of the scale was calculated and norms of the scale were developed. The final form of the Perceived Teacher Engagement Scale with three key dimensions and 22 scale items were administered on a randomly selected 579

secondary school students studying in 10th grade in the government schools under West Bengal Board of Secondary Education (WBBSE). The schools were situated in the seven districts (Cooch Behar, Purba Bardhaman, Nadia, Murshidabad, Birbhum, North Dinajpur and Hooghly) of West Bengal. After scoring (see Table 3.39) the 579 answer sheet, the statistical results obtained are as given in Table 3.40:

Table 3.39

Scoring system

Nature of item	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Positive	5	4	3	2	1
Negative	1	2	3	4	5

Table 3.40

Statistical results

Sl. No.	Sub-scales	N	M	SD
1	Perceived Cognitive-physical engagement	579	21.20	5.24
2	Perceived Socio-emotional engagement	579	25.04	6.11
3	Perceived Pedagogical engagement	579	21.84	5.20
	Overall scale	579	68.08	10.55

On the basis of the statistical results presented in Table 3.40, z-Score Norms dimension-wise and for the overall scale have been developed and presented as: Dimension-wise Norms in Table 3.41, Table 3.42 and Table 3.43 whereas for the overall scale in Table

3.44, respectively. Norms for interpretation the level of each dimension and overall Perceived Teacher Engagement Scale have been presented in Table 3.45.

Table 3.41

z-score norms for Cognitive-Physical Engagement sub-scale

M= 21.20 SD= 5.24 N= 579

Raw score	Z-score	Raw score	Z-score	Raw score	Z-score
10	-2.14	18	-0.61	26	+0.92
11	-1.95	19	-0.42	27	+1.11
12	-1.76	20	-0.23	28	+1.30
13	-1.56	21	-0.04	29	+1.49
14	-1.37	22	+0.15	30	+1.68
15	-1.18	23	+0.34	31	+1.87
16	-0.99	24	+0.53	32	+2.06
17	-0.80	25	+0.73	33	+2.25

Table 3.42

z-score norms for Socio-Emotional Engagement sub-scale

M= 25.04 SD= 6.11 N= 579

Raw score	Z-score	Raw score	Z-score	Raw score	Z-score
10	-2.46	20	-0.82	30	+0.81
11	-2.30	21	-0.66	31	+0.98
12	-2.13	22	-0.50	32	+1.14
13	-1.97	23	-0.33	33	+1.30
14	-1.81	24	-0.17	34	+1.47
15	-1.64	25	-0.01	35	+1.63
16	-1.48	26	+0.16	36	+1.79
17	-1.32	27	+0.32	37	+1.96
18	-1.15	28	+0.48	38	+2.12
19	-0.99	29	+0.65	39	+2.28

Table 3.43

z-score norms for Pedagogical Engagement sub-scale

M= 21.84		SD= 5.20			N= 579
Raw score	Z-score	Raw score	Z-score	Raw score	Z-score
9	-2.47	18	-0.74	27	+0.99
10	-2.28	19	-0.55	28	+1.18
11	-2.08	20	-0.35	29	+1.38
12	-1.89	21	-0.16	30	+1.57
13	-1.70	22	+0.03	31	+1.76
14	-1.51	23	+0.22	32	+1.95
15	-1.32	24	+0.42	33	+2.15
16	-1.12	25	+0.61	34	+2.34
17	-0.93	26	+0.80	35	+2.53

Table 3.44

z-score norms for overall Perceived Teacher Engagement Scale (PTES)

M= 68.08		SD= 10.55			N= 579
Raw score	Z-score	Raw score	Z-score	Raw score	Z-score
38	-2.85	58	-0.96	78	+0.94
39	-2.76	59	-0.86	79	+1.04
40	-2.66	60	-0.77	80	+1.13
41	-2.57	61	-0.67	81	+1.22
42	-2.47	62	-0.58	82	+1.32
43	-2.38	63	-0.48	83	+1.41
44	-2.28	64	-0.39	84	+1.51
45	-2.19	65	-0.29	85	+1.60
46	-2.09	66	-0.20	86	+1.70
47	-2.00	67	-0.10	87	+1.79

48	-1.90	68	-0.01	88	+1.89
49	-1.81	69	+0.09	89	+1.98
50	-1.71	70	+0.18	90	+2.08
51	-1.62	71	+0.28	91	+2.17
52	-1.52	72	+0.37	92	+2.27
53	-1.43	73	+0.47	93	+2.36
54	-1.33	74	+0.56	94	+2.46
55	-1.24	75	+0.66	95	+2.55
56	-1.15	76	+0.75	96	+2.65
57	-1.05	77	+0.85	97	+2.74

Table 3.45

Norms for interpretation of the levels of Perceived Teacher Engagement: Dimension-wise and overall scale

Sl. No.	Raw score range				z-score range	Levels of perceived teacher engagement
	Dimension-wise raw score range			Full scale		
	PCPE	PSEE	PPE			
1	32 & above	38 & above	33 & above	90 & above	+2.01 & above	Very high
2	28 to 31	33 to 37	29 to 32	82 to 89	+1.26 to +2.00	High
3	24 to 27	29 to 32	25 to 28	74 to 81	+0.51 to +1.25	Moderately high
4	19 to 23	22 to 28	20 to 24	63 to 73	-0.50 to +0.50	Average
5	15 to 18	18 to 21	16 to 19	55 to 62	-1.25 to -0.51	Moderately low
6	11 to 14	13 to 17	12 to 15	47 to 54	-2.00 to -1.26	Low
7	10 & below	12 & below	11 & below	46 & below	-2.01 & below	Very low

Note. PCPE= Perceived Cognitive-physical engagement, PSEE= Perceived Socio-emotional engagement, PPE= Perceived Pedagogical engagement

d. Factorial invariance of the student engagement measure across gender

Firstly, the unconstrained model (M_0) was evaluated to test the configural invariance of the 3-factor measurement model across the groups based on the demographic variables (i.e. gender, the locale of school, and school grade). Results (see Table 3.46) showed acceptable model fit indices indicating that the factorial structure stands invariant across all the compared groups. This model served as a reference for comparing the subsequent nested models with increasing restrictions. Secondly, the first constrained model i.e. Model 1 (M_1) (constrained by standardized factor loadings) was examined to check the equivalence in the matrix of standardized factor loadings of the different sub-samples (metric invariance). Results (see Table 3.46) showed satisfactory fit indices for M1 where no significant increment in RMSEA and CFI (ΔRMSEA<.015, ΔCFI<.01; Cheung and Rensvold, 2002) was recorded as compared to the baseline model (M_0). This indicated that there were no significant differences between the baseline model (M_0) and the constraint model (M_1). Therefore, there were no significant differences in the pattern of standardized factor loadings indicating that the factorial structure of the construct was equivalent across the groups. The results suggested that M_0 is equivalent to M_1. Hence, the metric invariance of the measurement model was established.

Thirdly, the second nested model (M_2) constrained by intercepts was evaluated. Results (see Table 3.46) showed acceptable model fits for Model 2 (M_2) and proposed equivalence of intercepts (scalar invariance) of the different groups as the model fit indices of M_2 were found similar to the previous constrained model (M_1). While comparing the fit indices, it was found that the difference between RMSEA and CFI values did not exceed the threshold limits. Therefore, the constraint model (M_2) was equivalent for all groups and the invariance of intercepts was supported.

Finally, the third constrained model i.e. Model 4 (M_4) was proposed where the error variances and covariances were restricted to be equal among the groups (residual invariance). When Model 4 (M_4) was compared with the previous constraint model (M_3), it was found that though the constraint model (M_4) showed acceptable fit indices, the difference in RMSEA and CFI exceeded the criterion values (ΔRMSEA= .015, ΔCFI= .01). Therefore, the constraint model did not fulfill the criteria of residual invariance and,

hence, error variances and covariances could not be considered equivalent across groups. However, the measurement model of student engagement satisfied 'strong invariance' (Byrne, 2008) across groups. Similarly, the hierarchical steps were followed for the other groups based on the locale of school and school grade. Results (see Table 3.46) show that the measurement model of student engagement followed strong invariance across gender.

Table 3.46

Evaluation of factorial invariance of the measurement model of teacher engagement

	χ^2	df	χ^2/df	$\Delta\chi^2$	Δdf	CFI	RMSEA [90% CIs]	ΔCFI	ΔRMSEA
First-order 3-factor teacher engagement measurement models									
M_0	588.09	412	1.43	---	---	.987	.027 [.022-.032]	---	---
M_1	614.80	434	1.42	26.71	22	.985	.027 [.022-.032]	.002	0
M_2	643.77	456	1.41	28.97	22	.983	.029 [.024-.031]	.002	.002
M_3	648.75	459	1.41	5.38	3	.983	.029 [.024-.031]	0	0
M_4	754.66	481	1.57	105.91	22	.963	.031 [.027-.035]	.046	.020

Note. M_0: fully unconstrained (Baseline) model [Configural invariance], M_1: measurement model with fixed standardized factor loadings [Metric invariance], M_2: measurement model with fixed standardized factor loadings and item intercepts [Scalar invariance], M_3: measurement model with fixed standardized factor loadings, item intercepts and factor covariances, M_4: measurement model with fixed standardized factor loadings, item intercepts, factor covariances, and error variances [Structural covariance invariance]; *p<.001

3.7 Limitations

The current study relied on the self-reported measures of student engagement. However, self-report measures are likely to contribute to social desirablity bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003) and hence, results in low validity of the self-report questionnaires (Heine, Lehman, Peng, & Greenholtz, 2002). Though, self-reports suffer

from response bias (McCroskey, Sallinen, Fayer, Richmond, & Barraclough, 1996), self-report instruments are still preferred as those are comparatively more cost-effective and time-saving. Perhaps, a multi-informant approach considering teachers' perceptions of students' engagement might be more meaningful. Besides, multiple methods of data collection like observation of the students regarding their engagement in classroom settings, interviewing of the students regarding their experiences during the classes might be useful for the study.

However, in the research design several precautionary measures were taken during tool construction and during the process of data collection which might out-weight the limitations discussed earlier. For example, the sample was selected exclusively using simple random sampling technique. The tools for measuring the study variables (viz. student engagement and perceived teacher engagement) were constructed following rigorous process (pre-testing, piloting, EFA and CFA of the measurement models) that resulted in high reliability of both tools. Further, both the tools were gender invariant that allowed the researcher to compare the mean scores of those variables across students' gender.